Washington University in St. Louis

Washington University Open Scholarship

Olin Business School Electronic Theses and Dissertations | Washington University Open Scholarship

Spring 5-15-2022

Essays in Behavioral Corporate Finance, Corporate Governance, and Higher Purpose

Kingsley Wabara
Washington University in St. Louis, k.wabara@wustl.edu

Follow this and additional works at: https://openscholarship.wustl.edu/olin_etds

Part of the Business Administration, Management, and Operations Commons, Economics Commons, and the Finance and Financial Management Commons

Recommended Citation

Wabara, Kingsley, "Essays in Behavioral Corporate Finance, Corporate Governance, and Higher Purpose" (2022). *Olin Business School Electronic Theses and Dissertations*. 8.
https://openscholarship.wustl.edu/olin_etds/8

This Dissertation is brought to you for free and open access by the Washington University Open Scholarship at Washington University Open Scholarship. It has been accepted for inclusion in Olin Business School Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST. LOUIS

Olin Business School

Dissertation Examination Committee:

Todd T. Milbourn, Chair

Marcus Berliant

Xing Huang

Mark Leary

Anjan V. Thakor

Essays in Behavioral Corporate Finance, Corporate Governance, and Higher Purpose

by

Kingsley W. Wabara

A dissertation presented to

The Graduate School of

Washington University in Saint Louis

in partial fulfilment of the

requirements for the degree of

Doctor of Philosophy

May 2022

Saint Louis, Missouri

© 2022, Kingsley Wabara

Table of Contents

List of Figures

List of Tables

Acknowledgment

I thank my dissertation committee – Todd T. Milbourn, Anjan V. Thakor, Mark Leary, Xing Huang, and Marcus Berliant – for their support and impeccable advice. Todd: for your motivation, big-brotherly encouragement, and permission to fully express my ideas, a gesture that, without an iota of doubt, exponentially accelerated my progress in the program and on the job market. Anjan: for being a remarkable teacher, a worthy example for research and service in academia, and a role model for competence and writing in Finance and Economics. Mark: for your exemplary gentility, expertise, and encouragement, and for demonstrably being my champion on the job market. Xing: for severally granting me audience and providing constructive feedback as I worked to develop the main ideas in the first chapter of this dissertation. Marcus (*"Uncle Marcus"*): for generously and faithfully calling us professors as though we already were, for being an inspirational teacher and for showing me a paper of yours that provides solid theoretical support for the main ideas in the second chapter.

I am also thankful to all students, staff, and other members of the faculty of the Olin Business School, especially Taylor Begley, Todd Gormley, Ohad Kadan, Asaf Manela, Radha Gopalan, Phil Dybvig, Jason Donaldson, Janis Skrastins, Jeremy Degenhart, Andrew Knight, Jessica Hatch, Brett Green, Armando Gomes, Guofu Zhou, Jennifer Dlugosz, Andy Neuhierl, and Tat Chan for their kind advice and valuable comments. I also thank Raghavendra Rau for his helpful comments and I am exceedingly grateful to the Wells Fargo Advisors Center for Finance and Accounting Research (WFA-CFAR) for the 2021 Outstanding Ph.D. Student Paper Award (received for my composite essay consisting of chapters 1 and 2 of this dissertation).

Finally, I am extraordinarily grateful to my entire family for their complete understanding.

Kingsley Wabara

Washington University in St. Louis

May 2022

To my amazing children – Nneoma (*"the French"*), Obim (*"the Texan"*), and Umechi (*"the Missourian"*) – for playing only little and working extremely hard to keep their "promise to make absolutely zero noise as dad was always busy at work in the home office," even though, to them, I was only using the laptop computer in the other room; and to my wife – Morayo – for again putting her career as a general practice physician on hold to focus on helping with the kids and the home, and for wilfully joining me on an audacious and unconventional journey to complete a full-time PhD program in Finance in the USA (after my dozen-plus years of accelerated career development in the industry).

To the loving memory of my grandmother – Isabella Wabara – who always encouraged me to seek the basic intuition behind every idea, study the scriptures deeply, never cease to strive to clothe myself with godly ambitions to make positive impacts in the world, and to forever remain focussed, disciplined, bold, humble, tenacious, forthright, generous, and quickly forgiving.

Above all, to the good Lord to whom all glory in heaven and on earth belongs, from whom comes every good and perfect gift and in whom there is no variableness (*Psalm 115:1, James, 1:17*).

ABSTRACT OF THE DISSERTATION

Essays in Behavioral Corporate Finance, Corporate Governance, and Higher Purpose

by

Kingsley W. Wabara

Doctor of Philosophy in Finance

Washington University in St. Louis, May 2022

Research Advisor: Professor Todd T. Milbourn

Does individual financial risk-taking behavior indeed vary by gender? If yes, how exactly? Moreover, what happens when individuals with differential tendencies (e.g., risk-taking, sense of higher purpose, et cetera) come together in small groups such as the corporate board to decide on financial risk-taking or conduct intense board oversight? What role does group or board structure play in all of these? In effect, how does the within-group power and influence asymmetry affect the manifestation of the individual tendencies in boards' decisions, corporate behavior, and financial outcomes? The lack of conclusive answers to these questions is at the core of several vital debates at the intersection of behavioral economics and corporate finance (corporate governance in particular) with significant implications for government policies and legislations.

For example, while practitioners have mostly positive anecdotes on the impact of board gender diversity on firms' financial performance, the empirical evidence in the academic literature remains mixed, with the debates still intensifying rather than abating.

Also, while the original intention of intense board oversight was to help nip corporate misconduct in the bud and preserve long-term firm value, the extant empirical evidence shows an unintended loss in firm value instead. Is this always the case? Specifically, are there any individual attributes of the monitoring-intensive directors that can alter their approach to intense board oversight and potentially help reverse this unintended trend?

This dissertation provides new perspectives and answers to these questions and extracts practical, generalizable, and broadly applicable intuitions therefrom. Across three essays, each of which constitutes a chapter of the dissertation, I use some combination of novel data, empirical or analytical tools, and even new methods to shed new light on the above subjects and debates. My primary goal is to contribute to the reconciliation of the various mixed or extant adverse evidence. The first chapter focuses on the relationship between gender and risk-taking behavior and how the within-group power and influence distributions affect the manifestation of such individual tendencies in small groups. Specifically, I use the multiple Emmy Award-winning financial decision-making TV game show, *Cash Cab*, as a pseudo-laboratory. I find that, on average, compared with a male, a female will take less excessive financial risks (i.e., mainly when the ex-ante expected profit from the financial risk-taking is non-positive). However, in small groups, personal power and influence, not just numerical strength, determine whether such individual tendencies (e.g., risk-taking appetites) manifest in the collective decisions or outcomes. A main and generalizable intuition from the pseudo-lab setting is as follows: power and influence guarantees voice and inclusion (and therefore impact) within-group.

The second chapter builds on the main intuition from the first but uses the corporate board as the primary setting. In this new context, I evaluate how the within-board power and influence asymmetry affects the impact of board gender diversity on firms' financial outcomes. To do this, I avoid the confounding effects of exogenous gender-diversification processes (such as gender quota legislations or laws). Nevertheless, I deal with the potential threat of endogenous selection using a sequence of stacked difference-in-differences estimations on mainly three sets of structured quasi-temporal event counterfactual samples. Overall, I find that if a female director is unlikely to have any personal power or influence on the board, her addition to the board will have no significant impact on firm risk-taking and performance. However, with increasing power/influence on the board (via greater numerical strength or non-token aggregate position), female directors will tend to reduce the excessive risk-taking behavior of the firm and, to the extent that the gender-diversification is non-disruptive, the expected risk-reduction effect can feature significant increases in profitability and firm value.

Finally, the third chapter turns to the unintended consequences of intense board oversight. Using a novel measure, this essay shows that when directors that are likely to have a higher sense of purpose lead the intense board oversight, the benefits improve significantly: earnings quality is higher; excess CEO compensation reduces; the time horizon perspective on CEO performance evaluation becomes longer.[1] Overall, the findings show that religious monitoring-intensive directors differentially influence intense board oversight results and, thereby, help infuse or propagate a corporate culture consistent with an authentic organizational higher purpose.

[1] A direct extension of this work outside of the scope of this dissertation further shows, both theoretically and empirically, that the longer-term perspective on CEO-performance evaluation is valuable.

Chapter 1

Group Gender Diversity, Individual Power and Influence, and Collective Risk-Taking

1.1. Introduction

> "While the individual [wo]man is an insoluble puzzle, in the aggregate [s]he becomes a mathematical certainty. You can, for example, never foretell what any one [wo]man will be up to, but you can say with precision what an average number will be up to. Individuals vary, but percentages remain constant. So says the statistician.[(a)]" "[A] time [comes when] the demographic changes … meet up with the … power that is possible. [(b)]" "Power reveals [who people truly are]. [(c)]"[2]

Does group gender diversity (really) matter for collective risk-taking? If yes, how exactly? Suppose the risk embedded in an investment decision is qualitatively excessive (e.g., the risk-taking opportunity per participant or group is a one-shot event, and the expected profit from the financial risk-taking is non-positive). How does individual financial risk-taking behavior vary by gender? What happens to the risk-taking behavior of small groups (of different gender compositions) when each member can ultimately exert the equivalent of absolute power or influence on the group's willingness to take

[2] (a) Arthur Conan Doyle, The Sign of the Four (1890); (b) Stacey Abrams on the Late Late Show with Stephen Colbert (November 2020); (c) Diane Coutu, Lessons in Power: Lyndon Johnson Revealed, HBR, April 2006.

significant financial risks? The lack of unambiguous fundamental answers (in the literature) to these questions parallels the mixed evidence on subjects such as the impact of board gender diversity on firms' risk-taking and performance.[3] Moreover, even if one assumed that a given gender type was generally more risk-averse than the other, it may still be unclear how such differential individual financial risk-taking tendencies would manifest in the collective financial decisions of small groups. Given the significant potential for individual power/influence asymmetry within-group, these ambiguities can be hard to resolve.

One approach to overcoming these challenges may require a single empirical setting in which (i) both individual and group risk-taking behaviors are observable and (ii) the impact of power and influence asymmetry is appropriately controlled for (or fixed as a feature). The *Cash Cab* is an excellent example of such a setting. *Cash Cab* is a multiple Emmy Award-winning television game show where unsuspecting people (individuals or groups) board a cab to head to a destination. Once they are in the cab, the game host (who also doubles as the cab driver) informs the participants that they are in a game show. If they agree to play, the participants work together, answering batches of general knowledge questions and, in return, earn discrete amounts of money for providing the correct answers to the general knowledge questions. The cash reward increases as the questions become more challenging. The participants[4] may use up to two external sources of help and are allowed a maximum of two errors (else, they forfeit their earnings and quit the cab prematurely). However, if the participants get to their destination successfully, providing

[3] For instance, although the literature on the relative risk aversion of males and females is substantial, its conclusions have not been entirely consistent: Jianakoplos and Bernasek (1998), using US sample data of household holdings of risky assets, conclude that single women are more financially risk-averse than single men. Their results are reinforced by, among others, Sunden and Surrette (1998), using data from the US Surveys of Consumer Finances; Eckel and Grossman (2002, 2008) using gambling data; and Stark and Zawojska (2015), theoretically. However, using survey data, Adams and Funk (2012) suggest that female directors are different from women in the general population and are more risk-loving than male directors.

[4] Figure 1.8 shows that the participants are vastly diverse.

mostly correct answers without exceeding the maximum number of errors, they earn their cumulative winnings.[5] Nevertheless, just before they disembark the cab, the host makes the participants a double-or-nothing offer for the chance to answer one more random (bonus) video question, a risky proposition the participants may freely accept or reject.

Ex-ante, the expected profit from the risky bet[6] is precisely zero. Hence, we call this point of the game the Risky Investment Decision Point (RIDP). The risk embedded in the offer at the RIDP also appears relatively high, qualitatively at least, given the one-shot nature of the offer, per group, and the zero ex-ante expected profit. Nevertheless, the (collective) financial risk-taking cannot proceed if any member strongly[7] opposes the bet. We call this property of the *Cash Cab* game setting the *implicit individual veto power*. Also, no member of a group can force the collective financial risk-taking upon the rest. The latter implies that the implicit veto power in the *Cash Cab* is strictly unidirectional (i.e., it can be used to stop, but never to compel, the risk-taking). These features of the game show make the *Cash Cab* RIDP a great setting to examine the questions posed earlier.[8]

In sum, this study uses the *Cash Cab* setting as a pseudo-laboratory to evaluate what happens to the financial risk-taking behavior of small groups (of different gender compositions) when any member can exert the equivalent of absolute but unidirectional power/influence on the group's willingness to take qualitatively high financial risks. We focus on the participants' individual or collective decision-making at the RIDP (where the ex-ante expected profit from the risky financial bet is zero). We take no initial position on

[5] Figure 1.1 shows the distribution of the earnings, X.
[6] The bet is equivalent to buying the lottery $\begin{pmatrix} 2X & 0 \\ q & (1-q) \end{pmatrix}$ for the price of X. Ex ante, $q = 0.5$.
[7] The overriding decision-making rule at the RIDP is unanimity since every member possesses this implicit veto power to stop the collective financial risk-taking and the game host also checks to make sure that no individual member is in total dissent. However, each group member reserves the right to use the veto power or not, such that the apparent outcome may, sometimes, seem like a consensus. The implicit individual veto power in the game setting ensures that the group dynamics remain egalitarian.
[8] We provide more details of the *Cash Cab* television game show in section 1.2.

whether one gender is more risk-averse than the other. However, we reason that a pseudo-experimental examination of the financial risk-taking of the game participants at the RIDP might provide some valuable insights. Concretely, we hypothesize that, because each member of a *Cash Cab* group equally possesses some unidirectional veto power to enforce less collective risk-taking, the presence of one individual of the more risk-averse gender type will, on average, significantly reduce a group's willingness to take financial risks. However the converse scenario for the presence of the less risk-averse gender type will, on average, not be observed in the data, primarily because the unidirectional veto power can be used to enforce less collective risk-taking but not otherwise. We use data from the original series[9] of the game show to empirically evaluate our hypotheses.

Our analyses of the *Cash Cab* pseudo-laboratory data reveal some fascinating results unique to the setting[10]. Starting with the one-person games, we find that an individual female participant is, on average, more risk-averse than an individual male participant. Surprisingly, looking at the two-, three- or four-person group participants, we ultimately show that, on average, the presence (or addition) of just one female in (to) a small (previously homogeneous male) group in the *Cash Cab* reduces the group's willingness to take the financial risk at the RIDP. However, if a group (of at least three persons) in the *Cash Cab* consists of one female, adding more females does not change the group's risk-taking behavior. The key drivers of these results are discernible from the main features of the setting. Once a group earns the cumulative sum $X > 0$, each group member can mentally compute their earnings per person (EPP). The implicit acknowledgment of the EPP and the role of unanimity affords individual members the equivalent of veto power. The latter ensures that groups never purchase the lottery even if just one member

[9] The original series of the *Cash Cab* game first aired in the USA between 2005 and 2013.
[10] Group sizes vary from one to as many as four participants.

insistently says no (or, simply, vetoes). More so, the game host generally checks to ensure that the group is collectively happy to take the financial risk, with no absolute dissent.

In essence, the homogeneous male groups in the *Cash Cab* take more qualitatively excessive financial risks because the average male appears more willing to take financial risks at higher individual earnings than the average female, all other things equal. Also, conditional on one female being in a *Cash Cab* group (of at least three persons), the presence (or addition) of more females seems, on average, irrelevant for group risk-taking because the first average female in the group (with implicit veto power) converts the group, figuratively, to an average female-centric group in terms of relative group risk preferences. The converse scenario for the addition of one male to, say, a previously homogenous female group is, as hypothesized, not observed in the data because the individual veto in the game setting can only be used to stop, but never to enforce, the collective risk-taking.

In the appendix, we provide references to publicly viewable video data to demonstrate that several alternative explanations or conjectures, other than the vital game features (e.g., the implicit individual veto power), are not essential factors for evaluating the collective financial risk-taking behavior of the *Cash Cab* groups. For example, one such conjecture could be that all males in the homogeneous male groups could be risk-loving while all males in the non-homogeneous groups could be, at best, as risk-averse as the female(s) in their respective groups. We show that this conjecture eventually leads to circular analysis and fails to explain our results consistently.[11]

Together with the unique properties of the *Cash Cab* as a pseudo-laboratory setting, our results imply that, in small groups, personal or a subgroup's share of the within-group power/influence distributions, not just numerical strength, determine

[11] We provide more details of the internal validity and robustness analyses in sections 5 and 6.

whether individual or the subgroup's tendencies become manifest in the collective outcome; exactly so for the effect of gender on collective financial risk-taking. Furthermore, because the personal veto power implicit in the *Cash Cab* setting guarantees that no individual participant or subgroup is token, we expand the definition of a token individual/subgroup to include those whose relative share of the within-group power and influence distributions is unlikely to enable the manifestation of their average characteristics or tendencies (e.g., risk preferences, deliberative mechanisms, et cetera) in the ultimate collective decisions.

The above definitional expansion is significant because Kanter (1977) popularized that proportions (i.e., the relative numbers of socially and culturally different people in a group) are critical in shaping interaction dynamics. She identified four group types based on varying proportional compositions and described "skewed" groups as containing a large preponderance of one type (the numerical "dominants") over another (the rare "tokens"). Nevertheless, our findings in the *Cash Cab* pseudo-laboratory suggest that proportions must lead to sufficient within-group power/influence to be meaningful for the eventual evasion of tokenism or the guarantee of an exit from a token status. Put differently, a low-proportion subgroup, or even just one individual, with an enormous share of the within-group power and influence distribution, can be "the dominant," not "the token."

This chapter makes several contributions to the literature. Our findings shed new light on the mixed extant evidence on the relationship between gender and individual financial risk-taking. For example, Jianakoplos and Bernasek (1998), using US sample data of household holdings of risky assets, conclude that single women are more financially risk-averse than single men. Using survey data, Adams and Funk (2012) conclude that female directors are different from women in the general population and are more risk-loving than male directors. Our results suggest that every female need not be more risk-averse than any male, but whenever the financial risk-taking is qualitatively excessive, females are likely to be more risk-averse. Our unique setting and novel empirical analyses highlight that the within-group power/influence distributions impact whether individual tendencies (such as relative risk-taking behavior) manifest in small groups.

Our findings also provide new insights into tokenism. Rhode and Packel (2014) discussed “tokenism/critical mass” as one “barrier/solution” pair. Tokenism refers to the broader argument about whether the appointment of only one or two female or minority directors on a board will significantly improve board decision-making. Kanter (1977) found that token members often encounter “social isolation, heightened visibility ... and pressure to adopt stereotyped roles.” Kramer et al. (2006) argued that “critical mass” is necessary to realize the benefits of diversity on corporate boards fully. However, our findings suggest that “critical mass” must lead to sufficient power/influence within-group to be meaningful for the eventual evasion of tokenism or the guarantee of an exit from a token status. In other words, a low-proportion subgroup, or even just one individual, with an enormous share of the within-group power and influence distribution, can be “the dominant,” not “the token” member of the group. The preceding is consistent with the minority relations literature, which suggests that power, privilege, and prestige are more important than numbers for understanding relations between dominant and subordinate groups (e.g., see Gittler, 1956; Noel, 1968; Yetman, 1985; Zimmer, 1988).

Our results contribute to the broader discussions on the potential barriers to reaping the maximum benefits of gender diversity. Take the corporate boards, for example. Using interview data, Creary et al. (2019) highlight that collegial and egalitarian boards are more likely to accept and integrate differences of opinion. The board members believe that their expertise and willingness to learn are recognized and incorporated into the board's work. Interestingly, that collegial culture in which diversity is believed to thrive is a crucial feature of our empirical setting (i.e., the *Cash Cab* pseudo-laboratory, wherein an egalitarian culture is guaranteed by the individual veto power or the high individual share of the within-group power and influence distribution, implicit in the game setting).

This chapter is related to the subset of literature in economics that uses television game shows and natural experiments to study risk aversion and decision-making under uncertainty (e.g., see Gertner, 1993; Metrick, 1995; Beetsma and Schotsman, 2001; Fullenkamp et al., 2003; Hartley, 2005; Andersen et al., 2008; Post et al., 2008; Keldenich and Klemm 2011; Bliss et al., 2012; Kelley and Lemke 2013). In particular, Bliss et al. (2012) studied decision-making and risk aversion in the *Cash Cab*. Adopting a methodology from Gertner (1993), they focussed mainly on estimating the coefficients of risk aversion for groups of different sizes and conclude that, on average, the collective risk aversion of a group increases with its cumulative earnings and decreases with its size. We focus on the effect of gender composition on risk-taking and financial performance and include both the cumulative earnings and group size variables as controls.

To the best of our knowledge, this chapter is the first to document and analyze the implications of the veto power constraint in the *Cash Cab* game setting. Our study is also related to the literature that uses experiments to examine risky decisions. For example, Bogan et al. (2013) used student subjects and hinted that the risk-seeking behavior of a team might not necessarily be increasing in the number of males on the team. We extend that finding by explicitly showing that on average, and conditional on each member or, at least, the female member(s) of the group possessing some semblance of veto power, risk aversion does not necessarily increase with the number of women in a group.

The rest of the chapter proceeds thus: Section 1.2 gives a detailed overview of the *Cash Cab* game setting and delineates our hypotheses for this study. Section 1.3 presents the pseudo-laboratory data, sample, variables, and summary statistics. Section 1.4 discusses our identification strategy and assumptions. Section 1.5 presents the results. Section 1.6 discusses and critically analyzes the internal validity and potential extensions. Section 1.7 concludes.

1.2. *Cash Cab* as a Pseudo-Laboratory

The *Cash Cab* setting has been a fascinating source of empirical data and pseudo-laboratory for research in economics and psychology. For example, Bliss et al. (2012) used the *Cash Cab* as a "natural laboratory" to estimate the coefficients of risk aversion for groups of different sizes. Other authors have also used various bits of the dataset to primarily analyze decision-making under uncertainty (e.g., see Keldenich and Klemm 2011; Kelley and Lemke 2013). This study uses the unique empirical setting that the game show provides to more cleanly analyze the relative risk-taking behavior of all major group types by gender composition, particularly in the presence of a unidirectional individual veto power constraint. We provide, in the next two subsections, a general overview of the *Cash Cab* TV game show and a brief description of the development of our hypotheses.

1.2.1. Overview of the Cash Cab TV game show

Cash Cab is fundamentally a financial decision-making game show. It won the Daytime Emmy Award for Outstanding Game Show in 2008, 2009, and 2010. The original series of the game aired on the Discovery Channel between 2005 and 2013. The show was revamped (with slight modifications) in 2017. While much of the old series continues to air on the Game Show Network (GSN), bits of the game—including the revived series—may be found elsewhere online (including YouTube and the official website of Bravo TV).

In a game of *Cash Cab*, unsuspecting individuals or groups board a cab (taxi) to head to a destination only to be startlingly informed by the cab driver (who also happens to be the game host) that they "are in the *Cash Cab*." If they agree to play the game, they get both a free (potentially full) taxi ride to their destination and the chance to answer some questions for monetary rewards. The game starts immediately following the elucidation of the game rules by the host. Participants answer batches of random questions and earn cash for getting the answers right. The questions get increasingly random (or harder), and the reward per question also increases systematically, batch by batch. Participants may only make a maximum of two errors to remain in the game. Else they disembark prematurely. The *Cash Cab* participants who ultimately get to their destinations while still legitimately in the cab are deemed winners and are entitled to their cumulative earnings, $X > 0$.

Just before the "winning" participants disembark, the game host makes them a "double-or-lose-all" offer, which they may either accept or reject. Participants double (or lose) their cumulative earnings at this critical decision point only if their answer to a random bonus video question is right (or wrong). The bonus question is random in that the expected ex-ante probability of success is simply ½—i.e., before the final video question is asked, a participating individual or group will either know the answer or not. Consequently, we dub this final decision point a risky investment decision point (or RIDP for short). The latter constitutes a unique point from which to directly analyze the relative risk-taking behavior of all *Cash Cab* participants—groups and individuals, alike. Moreover, at this stage, any member of a winning group may insist that the group not purchase the lottery. Hence, we characterize this feature as a unidirectional veto power (because while the collective risk-taking cannot happen if a member insistently says no, no individual member of a group, or subgroup, can force collective risk-taking upon the rest).

Cash Cab participants may be single persons or groups (of different sizes—i.e., two, three, four, or, in a few instances, five persons, each). They may also come in various key gender compositions (e.g., homogeneous male, majority male, gender-balanced, majority female, and homogeneous female). We base our determination of gender on the physical characteristics (e.g., face, voice, etcetera), appearance (e.g., clothing style, et cetera), and deductions from the group interactions. The respective subgroups are also observably highly diverse along multiple other dimensions such as age, race, ex-ante relationships, wealth, etc. Moreover, the games are set in some of the biggest and most diverse cities in the world; the cumulative winnings under risk are economically significant; the immediacy of the outcome of the risky bet is common information; the linkage of outcome to the actual decision-makers (who know that the future states of the world are random) is direct; the general settings and key elements of the game are strikingly similar to and are easy to replicate in a simple board investment decision model; and it is possible to consolidate pure gender effects, after controlling for other factors.

In several dimensions, the *Cash Cab* game setting can be compared to some real-world entities such as the corporate board. For instance, *Cash Cab* is fundamentally a cooperative enterprise in which small groups of various sizes work together to earn a cumulative sum, X, and in the end, must reach an often consensus[12] investment decision on whether to distribute and part with their earnings or bet the same in a risky (investment) offer. Some elements[13] of the game are also strikingly similar to a simple

[12] Again, unanimity is the overiding rule; consensus is only an apparent outcome.
[13] See section 1.3 for more details of the *Cash Cab* game setting.

board investment decision model, and the cumulative winnings under risk are economically significant. However, different from what typically obtains on such real-world entities, but particularly interesting for our empirical analyses, the *Cash Cab* game setting explicitly reinforces a collegial within-group power structure and social dynamics in which every group member's voice, power, and/or influence can be magnified in an egalitarian manner, more so against qualitatively high financial risk-taking. Again, the immediacy of the outcome of the risky bet is common information; the linkage of the decision outcome to the actual decision-makers is direct; the participants are vastly diverse, and it is easy to consolidate pure gender effects after directly controlling for other game factors, such as learning effect, group size, and idiosyncratic group abilities.

1.2.2. Hypotheses development

We take no initial position on whether a female is more risk-averse or vice-versa. However, we conjecture that if risk-taking tendencies truly vary by gender, on average, and suppose that the group-interaction setting is such that each member of every small group (of various gender compositions) possesses some unidirectional veto power to ultimately shift the collective financial risk-taking behavior toward less risk-taking, then a pseudo-experimental examination of many of such groups might help resolve the debate.

Specifically, we hypothesize that, because each member in a *Cash Cab* group equally possesses some unidirectional veto power to ultimately enforce less collective risk-taking, the presence of one individual of the more risk-averse gender type will, on average, significantly reduce a group's willingness to take financial risks. However, and especially because of this unidirectional usage of the veto power to enforce less collective risk-taking but not otherwise, the converse scenario for the presence of the less risk-averse gender type will, on average, not be observed in the data.

1.3. Data

From June 2018 to December 2018—i.e., approximately six calendar months—we studied (game by game and frame by frame) well over 12,000 minutes of paid subscription videos of the *Cash Cab* games available[14] to public subscribers on the Discovery Channel website. We hand-collected all relevant data for this study using paper and pen. Our sample dataset consists of nine (out of the eleven seasons) of the original series. The games in our sample appear to have first aired on television in the USA and perhaps elsewhere, on an approximately one season a year basis (that is, over a total period of about 9 years, between 2005 and 2013). A season of *Cash Cab* consists of about 40 episodes. An episode consists of three games that are completed in roughly 30 minutes (including commercials).

The winning groups earn, keep, or lose their respective cumulative earnings (cash) in just about 10 minutes. We directly collected multiple group performance data on about 1,100 participating individuals or groups (of various sizes and gender compositions). With an average of about three or more persons per group in our sample, we directly observed well over 3000 different individuals in the unique group interaction and financial decision-making pseudo-laboratory that *Cash Cab* provides. Eight (out of the nine seasons in our sample were set in the city of New York. The other season was set in the city of Chicago. Both cities are two of the biggest and most diverse cities in the USA (and the world, in general). The latter fact is reflected, at least observationally, in the racial appearance of a large proportion of the game groups. Although some random snippets of

[14] In December 2018, the Discovery Channel (DC) updated its website and, in the process, removed much of the original series (seasons 1 to 9, and 11; most of which we had already studied and the relevant data of which we had already hand-collected) leaving, as at the date of our last data collection on the website, just three seasons (10, 12, and 13), the last two of which belong to the new (i.e., the revamped) series. The Game Show Network (GSN) still airs some of the original series but does not appear to have tagged the games in the same sequential order. Consequently, any further references to a game episode on the GSN are made to only enhance easy access for viewing purposes. We restrict our data analysis to the (original series) data collected on the Discovery Channel website.

different episodes of the games may be found elsewhere[15] online, we have nonetheless, and for data integrity purposes, restricted our sample data collection to the original source.[16]

1.3.1. Cash Cab variables creation and sample statistics

We define the key variables for our empirical analyses in the *Cash Cab* pseudo-laboratory in Appendix 1.1 Table 1. The variables include group size, gender composition, external consultation history, error rate, cumulative earnings (X), season (or year), and decision (at the RIDP). See also Table 1.1 for details of the summary statistics for these variables. We provide, below, some brief descriptions of their sample statistics.

My main *Cash Cab* sample consists of a total of 1047 groups of different sizes and gender compositions. One-person group or, more appropriately, individual participants (male and female) account for about 12% of the data. Two-person groups account for about 43% of the data, while 3-person and 4-person groups account for about 27% and 18% of the data, respectively. We observed only a very sparse dataset on 5-person groups (i.e., less than or equal to 3 groups in total) and therefore excluded them from our main sample.

Expectedly, the gender subgroups also come in various sizes (ranging from one to four persons) and are highly diverse in multiple other dimensions such as age, ex-ante relationship, race, et cetera. About 28% of the groups in our sample are homogeneous male groups. Approximately 14% are majority male groups consisting of either 2 men and 1 woman (in the case of a 3-person group) or 3 men and 1 woman (in the case of a 4-person group). Approximately 28% are gender-balanced groups consisting of either 1-man / 1-woman (in the case of a 2-person group) or 2-men / 2-women (in the case of a 4-person

[15] Youtube, GSN TV, Amazon Prime, Hulu TV, etcetera.
[16] https://www.discovery.com/tv-shows/cash-cab/

group). About 11% are majority female groups consisting of either 2-women and 1-man (in the case of a 3-person group) or 3-women and 1-man (in the case of a 4-person group). About 20% of our main sample is homogeneous female groups (again, of various sizes).

Of the 1047 groups in our main sample, approximately 36% did not use any external help to complete the requisite game tasks before either exiting the game prematurely or successfully arriving at the RIDP. The remaining 64%, however, either used some assistance or used the maximum number of help allowed for each participant group. Also, of the 1047 groups in our sample, approximately 9% made no errors before successfully arriving at the RIDP. The remaining 91%, however, either made some errors before successfully arriving at the RIDP or exceeded the maximum number of errors allowed for each participant group and therefore exited the game prematurely, in which case they forfeited their cumulative earnings up to their point of exit. Specifically, about 37% of the 1047 groups exceeded the maximum number of errors and therefore exited the game prematurely, forfeiting their cumulative earnings up to their points of exit. The remaining 63% arrived at the RIDP with no errors, some errors but below the maximum permissible number of errors, or with exactly the maximum permissible number of errors.

An outsized proportion of the non-zero group cumulative earnings ($X > 0$) lies between the $400 and $1,200 interval. A small proportion lies below the $400 mark and an even smaller proportion lies above the $2,800 mark. Historically, and particularly for the original *Cash Cab* series to which our sample belongs, only a few groups have managed to obtain a cumulative earning exceeding $3,000 for the roughly 10-minute work. And as previously indicated, all sums won appeared highly economically significant to the set of participating individuals in our sample, observationally at least.

Of the 661 groups of various sizes and gender compositions in our main sample that successfully arrived at the RIDP, approximately 58% rejected the offer to purchase the lottery while the remaining 42% accepted the offer to purchase the lottery. The prior probability of success at the RIDP is, at least in expectation, not much better than 50%. For example, the only public information at the RIDP is that the marginal task required for the double-or-lose-all offer (i.e., the random bonus video bonus question) is "random."

1.4. Empirical Analyses

In this section, we discuss the slicing of the *Cash Cab* data, outline the empirical identification strategy, identification assumptions, and empirical specification.

1.4.1. Cash Cab data slicing

To provide a background to the main intuitions that underpin our identification strategy in the pseudo-natural laboratory, We begin by slicing the data to expose some important details that may not be apparent otherwise. We show, in Figure 1.1, the frequency distribution of the cumulative group earnings together with that of the EPP. The EPP is computed by dividing the group earnings by the group size. Almost the entire non-zero cumulative earnings (i.e., the cumulative earnings of all groups that arrived at the RIDP) lie between the $250 and $2,000 interval. A small proportion lies above the $2,000 mark. The EPP distribution is also skewed to the right with much of the entire distribution lying in the interval between $100 and $1,000. As previously indicated, all sums are earned in just about 10 minutes and appear highly economically significant to the game participants.

We show, in Figure 1.2, the proportions of the participant groups that purchase the lottery upon arrival at the RIDP. The trend line indicates a linear decrease in the proportions from about 50% in season 3 to around 40% almost 9 years later. We also show, in Figure 1.3, the proportions of the participant groups in our main sample that purchase the lottery with the trend lines activated by group size. The trend lines indicate the presence of time effects and also that, on average, more 3 or 4-person groups arriving at the RIDP purchase the lottery than do the 2-person groups. The latter differences appear somewhat consistent across seasons. This pattern of increasing tendency to purchase the lottery with group size is also observed between the 3 and 4-person groups, albeit less markedly.

We show, in Figure 1.4, the proportions of the participant groups in our sample (with EPPs equal to or below a given threshold) that purchase the lottery upon arrival at the RIDP. We create this graph as follows. For each threshold, we collect the set of all groups with EPP less than or equal to that threshold and then compute the proportion of groups in that subset that purchase the lottery. The flat line portion of the graph indicates that the average lottery purchase rate, conditional on arrival at the RIDP converges to 42% for the sample (a trend we show in Figure 1.2 to be decreasing across time). The graph also reveals some combined earnings and group size effects on the decision to purchase the lottery that appear to exacerbate almost linearly below an EPP of $400.

We show, in Figure 1.5, the proportions of the participant groups in our sample that purchase the lottery upon arrival at the RIDP but, here, we compute the proportions for only 1-person groups or, more appropriately, the individual participants. The trend lines indicate the earlier discussed time effect and, with group size now fixed at 1, we observe a time-consistent average difference in risk-taking at the RIDP between an individual male participant and an individual female participant. On average, individual males appear to want to purchase the lottery more often than do individual females. Put differently, the individual females appear more risk-averse at the critical decision point.

We show, in Figure 1.6, the proportions of the participant groups in our sample that purchase the lottery upon arrival at the RIDP but with the data now split in two—one set comprising of all groups that did not use any external consultation and the other comprising of all groups arriving at the RIDP that did use some external consultation. The trend lines appear to indicate that the groups that did use some external consultation tend to purchase the lottery less often than those that did not use any external consultation and the differences appear to widen with time. Similarly, we show in Figure 1.7 the proportions of the participant groups in our sample that purchase the lottery but with one set comprising of all groups that had zero errors and the other comprising of all groups that had some errors. The trend lines appear to indicate, albeit less markedly, that the groups that had some errors tend to purchase the lottery more often than those that had zero errors. The differences here, in fact, appear to also widen with time, but appear much smaller compared to the more obvious external consultation effect.

1.4.2. Identification strategy and assumptions

The data splicing above helps to elucidate some fundamental intuitions that undergird our identification strategy. For instance, because it would be natural for every member of a *Cash Cab* group to care about their EPPs—before effectively pooling same for the collective risk-taking or declining to do so—we conjecture that groups comprising of more persons or those arriving at the RIDP with smaller group cumulative earnings ($X > 0$) will be more likely to bet their entire earnings for the chance to play the lottery. Interestingly, this conjecture is consistent with previous research findings using the *Cash Cab* data. Hence, we include both group size and amount won in our vector of controls.

Notwithstanding the above, clearly identifying the direct impact of gender composition on group risk-taking in the *Cash Cab* pseudo-laboratory setting will require further controls. For example, each participating group in the *Cash Cab* will naturally have its idiosyncratic task performance ability (which may be linearly or non-linearly derived from the individual task performance abilities of the respective group members). Moreover, the groups have different years of participation (which might affect their level of familiarity with the game). The groups also have different raw gender headcounts (and other more dispersed attributes such as age, race, wealth, et cetera) that might contribute to their absolute risk-taking behavior. More so, the idiosyncratic task performance ability (or the task-type-specific competency of each of the groups) is not directly observable.

To alleviate much of these identification challenges, we take advantage of the diversity of the game participants. Specifically, and as Figure 1.8 schematizes, the *Cash Cab* participants are vastly diverse in demographic characteristics such as age, gender, race, status, etc. Hence, although the choice to participate in the game is not a random assignment by the researcher (us), the deep diversity of the participants allows us to regroup them along progressive shades of gender categories—e.g., Homogenous Male, Majority Male, Gender-Balanced, Majority Female, and Homogenous Female. These pooled groups (with similar gender categories but multiple opposite characteristics) permit us to compute the average impact of gender diversity on risk-taking by re-enforcing the pure gender effects while averaging out other subgroup characteristics. Precisely, sorting the groups into broad gender categories helps to (i) consolidate the pure gender effects while, at the same time, average out much of the more dispersed attributes that may not have been explicitly measured and (ii) focus the general empirical analysis on the relative differences in risk-taking between the gender subgroups.

Next, because the unobservable idiosyncratic task performance ability of each group contributes directly to the group's error rate, tendency to consult externally, and amount won, we construct binary variables from the latter observable characteristics to account for the former. The group size and the year of participation in the game are directly observable, hence we particularly include them as controls to account for the fixed effects associated with the size of each group and the potential viewer learning effects that may have occurred across seasons since the game first aired on Cable TV in the USA.

Our identification assumption, therefore, is as follows. After controlling for the game season, group size, and idiosyncratic task performance abilities (i.e., a combination of the error rate, external consultation history, and amount won) and after sorting the raw groups into the broad gender categories—homogeneous male, majority male, gender-balanced, majority female, and homogeneous female—any residual differences in risk-taking between the gender subgroups are, on average, caused by their differences in consolidated gender subgroup types. See Figure 1.9 for a detailed schematic description of our identification strategy and assumptions in the *Cash Cab* pseudo-laboratory setting.

1.4.3. Empirical specification

The main dependent variable (i.e., the decision to / not to purchase the lottery at the RIDP) is an indicator. The analyses from the preceding subsections make clear that, in addition to gender composition, other (control) variables—e.g., season (year), group size, cumulative earnings, external consultation history, and the error rate—contribute in various ways to the explanation of the decision to purchase the lottery at the RIDP or not. Consequently, we choose the standard multiple logistic probability model for the empirical specification. For example:

$$Prob\left(Decision_j = Yes \mid \boldsymbol{IndVar_j}; \boldsymbol{\beta}\right) = \frac{1}{1+e^{-\beta^T IndVar_j}} \qquad (1)$$

Where $Prob\left(Decision_j = Yes \mid \boldsymbol{IndVar_j}; \boldsymbol{\beta}\right)$ is the conditional probability that the j^{th} group at the *Cash Cab* RIDP decides to purchase the lottery; $\boldsymbol{IndVar_j}$ is a vector including the main independent variable (i.e., 5-category gender composition variable) and the control variables (i.e., season, group size, cumulative earnings, external consultation history, and error rate); and $\boldsymbol{\beta}$ is a vector of the parameters that we wish to learn from the model. Put more intuitively, we specifically compute the following regression estimates:

$$\ln\left(\frac{\hat{p}_j}{1-\hat{p}_j}\right) = \hat{b}_0 + \sum_{i=1}^{n} \hat{b}_i IndVar_{ji} \qquad (2)$$

Where $\hat{p}_j$ is the estimated probability that the j^{th} participant group in our sample accepts the offer to purchase the lottery at the RIDP; $\hat{b}_0$ is a constant that represents the intercept from equation 1 (above); $IndVar_{j1}$ is the main explanatory variable while the other $IndVar_{ji}$s are the control variables; $\left(\frac{\hat{p}_j}{1-\hat{p}_j}\right)$ is the estimated odds ratio (or simply odds) of the j^{th} group deciding to purchase the lottery at the RIDP; and n is the total number of both the explanatory and the control variables. In other words, the left-hand side of equation (2) is simply the estimated natural logarithm of the odds of a participant group purchasing the lottery conditional on successfully arriving at the RIDP. We systematically build two regression models: Model 1 without, and Model 2 with, the error rate variable.

1.5. Results

In the subsections that follow, we present and critically analyze the results.

1.5.1. Main results

To fix ideas, the empirical results that follow are all, fundamentally, average effects, ceteris paribus. Notably, as Arthur Conan Doyle (The Sign of the Four, 1890) succinctly said: "While the individual [wo]man is an insoluble puzzle, in the aggregate [s]he becomes a mathematical certainty. You can, for example, never foretell what any one [wo]man will be up to, but you can say with precision what an average number will be up to. Individuals vary, but percentages remain constant. So says the statistician."

1.5.1.1. Gender composition and relative risk-taking

We show the main regression results in Table 1.3. After controlling for the season (or year) fixed effect, group size, cumulative earnings, and the external consultation history, we find that a homogeneous female group or any gender diverse group—be it majority male, gender-balanced, or majority-female group—is significantly less likely than a homogeneous male group to purchase the lottery at the RIDP. These results are significant for each of the groups at 5%, 5%, 1%, and 5%, respectively, and are unchanged with and without controlling for the group error rates. This implies that conditional on controlling for all other effects, participants' error rates do not matter and all non-homogeneous male groups are, on average, respectively more risk-averse than a homogeneous male group.

1.5.1.2. Non-linear gender diversification effect and hypotheses testing

For pictorial clarity, we plot in Figure 1.10 the differential regression coefficients for the gender subgroups (with the homogeneous male group as the reference group). This leads to the basic interpretation that all gender diverse groups—be they majority male, gender-balanced, or majority female—are each potentially less likely than (or at least as likely as) a homogeneous female group to purchase the lottery at the RIDP. This result suggests some non-linear (most likely concave) gender diversification effect on the groups' collective risk aversion. Consequently, using Model 1 as the reference multiple logistic regression model, we conduct a series of tests to evaluate the following set of null hypotheses:

- With the all-male group as the reference group, the differential effects of each of the diverse subgroups are not individually and significantly different from that of an all-female group

Appendix 1.1 Table 2 reports the p-values for the hypotheses tests. The All-Male column reproduces the actual p-values delineated in Model 1 with the standard asterisks (all of which are significant). The other four columns report the actual results of the tests of the null hypotheses. In all, we fail to reject the respective null hypotheses since the reported p-values are respectively greater than the significance level of 5%. Consequently, we update the plot of the differential regression coefficients of the gender subgroups in Figure 1.11 (still with the homogeneous male group as the reference group).

Again, a basic look at the updated plot leads to the interpretation that a homogeneous female group will be financially more risk-averse than a homogeneous male group, all other things equal. Also, any gender-diverse group—be it majority female, gender-balanced, or majority male—will produce a level of increased risk-aversion relative to that of a homogeneous male group that is at least as high as that which is likely to be produced by a homogeneous female group, all other things equal. It appears, however, that locked within this picture of an acute concavity of the gender-diversification effect is a more fundamental and analytical interpretation of the results. The latter relies hugely on the critical examination and full integration of the key features of the game setting.

1.5.2. A critical analysis of the acute concavity of the gender-diversification effect

First, we recap our results thus far. Looking at the individual participants, we find that on average, an individual female participant is more financially risk-averse than an individual male participant. This result is consistent with some extant analyses comparing relative risk-taking between men and women (e.g., Jianakoplos and Bernasek, 1998). Surprisingly, however, looking at the groups, we find that a simple gender-diverse group—be it majority female, gender-balanced, or majority male—will produce a level of increased risk-aversion relative to a homogeneous male group that is at least as high as that which is likely to be produced by a homogeneous female group, all other things equal. Below, we analyze the key driver for this result and other pertinent elements of our findings.

1.5.2.1. Key driver for the main results

Interestingly, the drivers of the key results are easily discernible from the observed collective decision-making conditions in the *Cash Cab* setting. Specifically, once a group earns the cumulative sum X > 0, each member of the group mentally computes his or her earnings per person (EPP). The implicit acknowledgment of the EPP arrogates to each member the equivalent of a veto power such that groups never purchase the lottery even if only one member expresses a serious reservation (i.e., insistently says no). More so, the game host checks to ensure that the group is happy to take the risk collectively and that there is no absolute dissent from any member of the group.

This individual veto power implicit in the *Cash Cab* game setting ensures that no participant is a token group member. Hence, we re-define a token member of a group as one whose share of the within-group power distribution is unlikely to enable the manifestation of their personal (risk) preferences in the group's collective (investment) decisions. In effect, the homogeneous male groups in the *Cash Cab* take more financial risks because the average male appears more willing to take financial risks at higher EPPs than does the average female. Moreover, conditional on one female being in a *Cash Cab* group, having more females seems, on average, irrelevant for group risk-taking because the first average woman in the group (with implicit veto power) converts it, figuratively, to an average female-centric group in terms of relative group risk preferences.

1.5.2.2. Does group size always matter for risk-taking?

Group size does not always matter for risk-taking. For example, looking at the group size control variable, we find that both the 3-person groups and the 4-person groups are each significantly more likely to purchase the lottery than the 2-person groups (see Table 1.4). However, conducting hypotheses tests on the group size fixed effects, we find that the average difference in risk-taking at the RIDP between a 3-person group and a 4-person group in the *Cash Cab*, all other things equal, is not statistically significant. This result also connotes, at least, an acute concavity of group risk aversion with increasing group size.

1.5.2.3. Is it just the presence (or also the addition) of one female?

The direct implication of the concavity of the group size effect is that, on average, there is no statistically significant difference in risk-taking between 3-person groups in the *Cash Cab* consisting of 2-men and 1-woman (or 1-man and 2-women) and 4-person groups in the *Cash Cab* consisting of 2-men and 2-women, ceteris paribus. It follows, therefore, that conditional on a non-homogeneous group having at least three persons, adding one more male or female does not significantly change the risk-taking behavior of the group.

1.5.3. Combining the main results and the critical analyses

The key driver for the main results in the *Cash Cab* is the implicit individual veto power (or the high individual share of the within-group power and influence distribution). The converse scenario for the addition of one male to, say, a previously homogenous female group is, as hypothesized, not observed in the data because the individual veto in the game setting can only be used to stop, and never to enforce, the collective risk-taking. The group size effect is also insignificant beyond three persons. Consequently, we conclude that, on average, adding one influential woman to a small previously homogeneous male group significantly reduces the group's willingness to take financial risks. If, however, the group (of at least three persons) consists of one such influential woman, adding more women does not significantly alter the risk-taking behavior of the group. This conclusion has some important connotations. For example, it suggests that group structure—particularly, the within-group power and influence distributions—is crucial for determining whether individual tendencies toward risk-taking become manifest in small group settings.

In other words, the addition of several non-influential females may not be nearly as impactful as the addition of one influential female to a previously homogeneous male group, except the increase in the proportion of females translates, non-disruptively, to more within-group power and influence for them. This is an important result, especially because previous research had popularized the necessity of a numerical threshold proportion for a minority subgroup to evade tokenism or exit a token status (i.e., to begin to have any meaningful impact within a larger group; e.g., see Kanter, 1977). While this may be valid in several situations, our findings in the *Cash Cab* pseudo-laboratory imply that greater proportions must lead to sufficient power and influence to be meaningful for the eventual evasion of tokenism or exit from a token status. Put differently, a low-proportion subgroup, or even just one individual, with an enormous share of the within-group power and influence distribution, can be "the dominant", not "the token" member of the group.

1.5.4. Internal validity

Explicit collegiality is, observably, a key feature of the *Cash Cab* setting and is guaranteed by the game feature of implicit individual veto power (i.e., the high individual share of the within-group power and influence distribution, which consistently explains our findings in the *Cash Cab* pseudo-laboratory). In this section, nevertheless, we outline some alternative conjectures or concerns that may be advanced as possible explanations for our results and then proactively explain or provide evidence as to why they do not apply in the *Cash Cab* setting and, therefore, neither significantly impact nor drive our results.

1.5.4.1. Do groups self-select by relative risk aversion?

One concern that might be advanced is the possibility that the men in the homogeneous male groups are all risk-loving while the men in the non-homogeneous male groups are, at least, as risk-averse as the woman or women in their groups. This conjecture appears wholesomely unnatural for the "unsuspecting" groups of persons that board the *Cash Cab* on the streets of New York and Chicago. It also clearly ignores the diversity (in multiple other dimensions) of the *Cash Cab* participants—some of whom are friends, colleagues, mere acquaintances, couples, or have other forms of family relationships.

Moreover, while this conjecture ex-ante concedes that women may be more risk-averse than men, it appears to do so rather extremely. For example, for this conjecture to explain why there is no significant difference in risk-taking, on average, between a majority male group of four consisting of 3-men and 1-woman and a homogeneous female group of four, one must further believe that there is some systematic natural rule that ensures that whenever people (who are diverse in multiple other dimensions) get together, it must be that only similarly risk-averse men allow a woman into their group and that the woman so allowed into the group may not be more risk-loving than the men in some homogeneous male group. We can structure this analysis in other extreme ways. However, to completely negate this conjecture, we present (in Appendix 1, Table 3) a descriptive reference to a sample of some easily accessible video data to show that the phenomenon implied by the conjecture under consideration is not an essential feature of the *Cash Cab* game setting and, therefore, does not drive our results. Again, the preceding conjecture is not an important (if at all a) feature of the *Cash Cab* pseudo-laboratory setting. Hence it does, on its face alone, fail to explain our results consistently.

1.5.4.2. Do groups self-select by relative wealth status?

Similarly, another concern that might be advanced is that it may be that the groups that accept to purchase the lottery at the RIDP are wealthier than the groups that do not. However, for this conjecture to explain our results, some difficult assumptions will need to be made. One, it must be that homogeneous male groups on either the streets of New York or the streets of Chicago waiting unsuspectingly to board a taxi must be systematically wealthier than their homogeneous female counterparts. Two, it must also be that the presence of a woman in a predominantly male group must lead to lower average group wealth relative to that of an equivalent homogeneous male group, all other things equal. It must also be that this lower relative wealth must be the key motivation for the higher level of risk aversion for the gender-diverse group. If, however, the latter is true, then it must also be that the woman strictly influenced the predominantly male group to which she belonged against taking the bet. Even so, the primary driver for the collective group decision must be the woman's influence on the group, albeit motivated by her poverty. This conjecture spirals away from itself and, therefore, does not drive our result.

1.5.4.3. Are our results in the Cash Cab driven by Room Effects?

Castillo, et al. (2013) studied how "who is in the room affects behavior directly absent some mechanism that makes others' decisions informative or payoff relevant." They implemented economic experiments that randomly varied the company in which a decision is made while shutting off any payoff or information channels. They found that the composition of the room in which individual decisions are made alters individual behavior. The effect they found was large and systematic. Women were willing to choose more risky alternatives when the proportion of men in the room increased. When women were surrounded by men, their behavior was statistically indistinguishable from men's. Women in that situation held double the amount in risky assets than women surrounded by other women. Hence, they conclude that the mechanism behind this change in behavior may be due to attempts to imitate the home-grown expectations of others.

While Castillo, et al. (2013) suggest that women are individually more financially risk-averse than men, their findings also appear to provide additional support that room effects do not drive our results for the *Cash Cab* groups. For example, if our results for the *Cash Cab* groups were driven by the room effects, then we should find that the majority-male groups are, on average, indistinguishable from the homogenous male groups and, in the same vein, distinguishable from the majority female groups. We find the contrary, specifically because the use of the veto power in the *Cash Cab* setting is unidirectional.

1.6. General Discussion

In this section, we anticipate/address some residual internal validity concerns that may be advanced, perhaps not as the primary drivers for our results but as potential sources of bias. Altogether, we outline the internal consistency of our findings. Finally, we suggest that our results could apply to the corporate board setting.

1.6.1. Integrating our key findings

Overall, we use hand-collected data from over 12,000 minutes of the original series of the television game show, *Cash Cab*. We find that, on average, the presence (or addition) of one influential female in (or to) a small previously homogeneous male group significantly reduces the group's willingness to take financial risks. If, however, a group (of at least three persons) consists of one such female, adding more females does not significantly alter the risk-taking behavior of the group. These findings and the key features of the *Cash Cab* setting suggest that the within-group power/influence distributions impact whether gender tendencies become manifest in small group settings.

1.6.2. Addressing non-major residual internal concerns

Some non-major concerns may remain, especially for those who are not familiar with the *Cash Cab* game setting and/or those who may still be worried about the usual endogeneity concerns when working with board data. For example, someone might want to ask: do the game producers air only interesting games? Will the results differ if the group members were randomly assigned? Are the *Cash Cab* results externally valid? For example, can the results be generally applicable to the corporate board setting? We address these questions in the subsections below.

1.6.2.1. Do the Cash Cab producers air only "interesting" games?

A couple of points help to mitigate and/or address this concern. One, given our identification strategy (in which the groups—diverse in multiple other dimensions—are sorted into key gender groups to consolidate pure gender effects), it is highly unlikely that any definition of "interesting" will materially affect our results. Two, such game pre-selections based on whether games are "interesting" do not inform the producers' decision to air the game episodes (e.g., see online google searches about the *Cash Cab*).

1.6.2.2. Will the results differ if the group members were randomly assigned by us?

Another concern that may still be advanced is that it might be better to replicate the *Cash Cab* game rules in a pure experimental setting (perhaps using student subjects) in which the respective group members might be more randomly assigned. We address this concern with our identification strategy. Moreover, our results are consistent with those hinted by a version of the kind of experiment that this concern appears to advance. For instance, Bogan et al. (2013), using student subjects, hinted that the risk-seeking behavior of a team might not necessarily be increasing in the number of males on the team. We confirm and then extend that finding by explicitly showing that on average, and conditional on each member—or, at least, the female member(s) of the group—possessing the equivalent of veto power, risk aversion does not necessarily increase with the number of women in a group. Our pseudo-laboratory study has the added advantage of explicit and repeated observations of the group dynamics in the *Cash Cab* using video data[17].

1.6.2.3. Can our results be generally applicable to a setting like the corporate board?

Consider that Bernile et al. (2018)—while using a composite diversity index and an exogenous source of variation in board gender diversity to primarily argue for the primacy of broader diversity beyond gender itself, nevertheless—documented that no single component (of their composite index) alone, except gender diversity, had a statistically marginal effect on return volatility (i.e., their proxy for risk-taking). The direction of the marginal effect of gender diversity on risk-taking that they found is consistent with our statistically significant findings in the *Cash Cab* pseudo-laboratory setting (where the link between the decision-makers and the risk-taking outcome is direct and we effectively consolidate pure gender effects, after controlling for other game factors).

[17] See Appendix 1.1 Table 3 for some references to easily accessible *Cash Cab* video data

Consider also that Faccio et al. (2016)—using data consisting of both publicly quoted and private European firms—documented that “transitions from male to female CEOs (or vice versa) are associated with economically and statistically significant reductions (increases) in corporate-risk-taking.” They also demonstrated that their “results are robust to controlling for the endogenous matching between firms and CEOs using a variety of econometric techniques.” Our findings in the *Cash Cab* pseudo-laboratory relate more specifically to when the financial risk-taking is qualitatively excessive (i.e., the ex-ante expected profit is non-positive). It raises the marginal question of whether females would simplistically reduce risk or reduce “excessive” risk.

Consider further that Creary et al. (2019) “interviewed nineteen board directors (fifteen women and four men) to learn whether and how corporate boards were benefiting from diversity. Combined, the board members held seats on forty-seven corporate boards in the U.S. across a variety of industries.” In sum, the interviewees surmised that “diversity doesn’t guarantee a better performing board and firm; rather, the culture of the board is what can affect how well diverse boards perform their duties and oversee their firms.” They also surmised that “in contrast to hierarchical boards, more egalitarian boards have a more ‘collegial’ board culture. The interviewees who felt that their boards reflected this quality explained how all board members were able to speak openly and ask questions at meetings and, in doing so, they felt that all opinions were respected.” They further surmised that “collegial boards are more likely to accept and integrate differences of opinion. Members of these boards believe that both their expertise and willingness to learn is recognized and incorporated into the board’s work.” Remarkably, such collegial culture in which diversity on the board is surmised to thrive is a crucial feature of the *Cash Cab* setting (wherein the egalitarian culture is guaranteed by the individual veto power or the high individual share of the within-group power and influence distribution, implicit in the game setting).

Our findings in the *Cash Cab* pseudo-laboratory also reveal that the within-group power/influence distributions impact whether individual tendencies, such as gender-driven risk preferences, become manifest in small group settings. Milbourn and Wabara (2021), referencing the main revelation from this chapter in their study of the impact of religiosity on intense board monitoring, show that the evidence that religious monitoring-intensive directors tend to further reduce excess total CEO compensation becomes highly significant when the lead independent director and/or a majority of the principal monitoring committee chairs are also religious (a condition that increases the within-board share of power/influence for the religious-intense monitors).

Overall, it might be helpful to borrow a leaf from this study to investigate several unanswered questions for the corporate setting. For example, it would be interesting to specifically study the ways in which the within-board power and influence asymmetries modulate the impact of gender diversity on firm performance and value. It would also be helpful to, in the process, provide some empirical evidence on whether female directors on the board would simplistically reduce risk or would reduce excessive risk. Although sufficient data availability might be a challenging factor, conducting this new research for publicly listed firms in the USA could be beneficial.

We take up these challenges in a subsequent study.

1.7. Conclusion

We use hand-collected data from over 12,000 minutes of the original series of the Emmy Award-winning television game show, *Cash Cab*, to examine what happens to the risk-taking behavior of small groups of different gender compositions when each member can ultimately exert the equivalent of absolute power and influence on the group's willingness to take significant financial risks. We find that, on average, the presence (or addition) of one influential female in (or to) a small previously homogeneous male group significantly reduces the group's willingness to take financial risks. If, however, a group (of at least three persons) consists of one such female, adding more females does not significantly alter the risk-taking behavior of the group. These findings and the key features of the *Cash Cab* setting suggest that the within-group power/influence distributions impact whether gender tendencies manifest in small group settings.

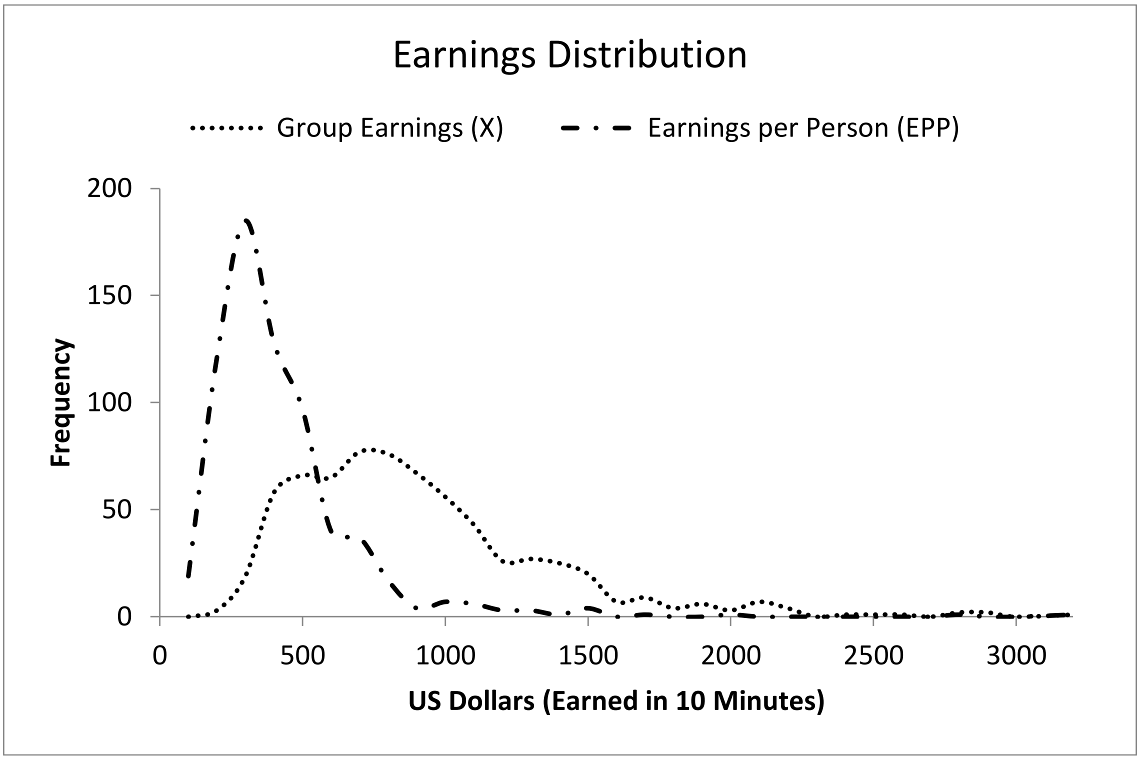


Figure 1.1
Group earnings and earnings per person (EPP) distribution
This figure plots the frequency distribution of the cumulative group earnings together with that of the EPP of the participant groups in our sample that arrive at the RIDP. The RIDP concept is described in the main text. The frequencies are based on 63.1% of the total 1047 groups in our main sample. See Table 1.1 for more details on the summary statistics. See Appendix 1.1 Table 1 for an outline of our main variables and the unique source of data. The EPP is computed by simply dividing the group earnings by the group size. All sums are earned in just about 10 minutes or one-sixth of an hour.

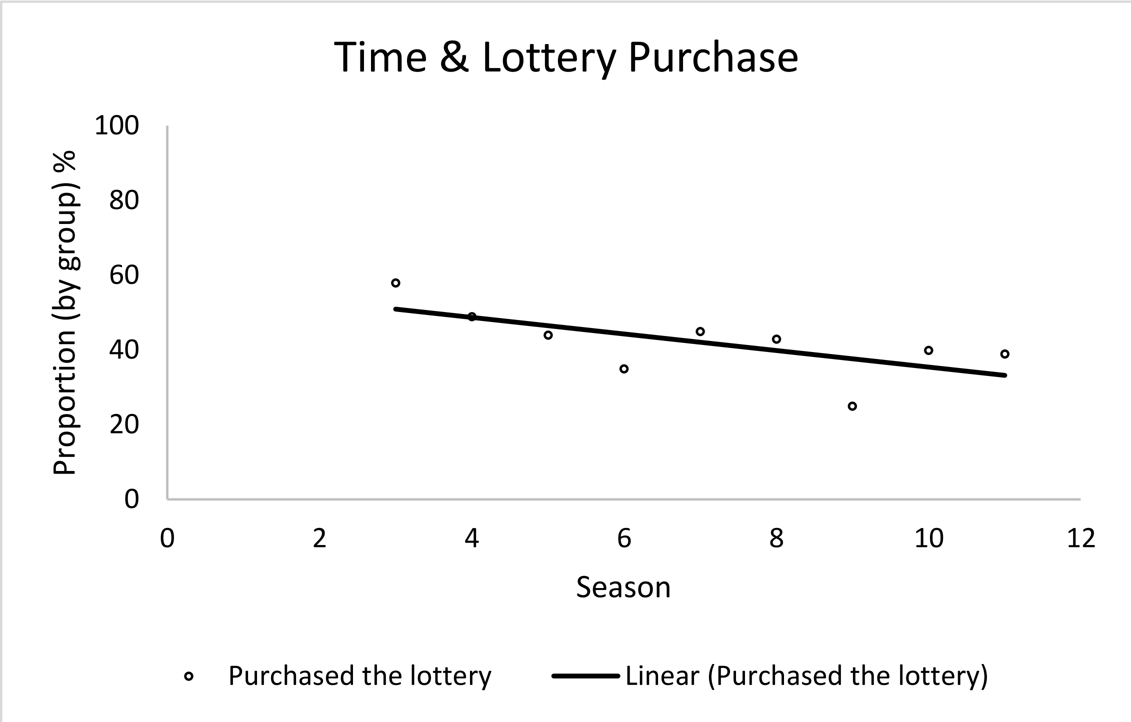


Figure 1.2
Time effect on lottery purchase at the RIDP
This figure plots the proportions of the participant groups in our sample that purchase the lottery upon arrival at the RIDP. The RIDP concept is described in the main text. The proportions are based on 63.1% of the total 1047 groups in our main sample. See Table 1.1 for more details on the summary statistics. See Appendix 1.1 Table 1 for an outline of our main variables and the unique source of data. The trend line indicates a linear decrease in the proportions from about 50% in season 3 to around 40% almost 9 years later.

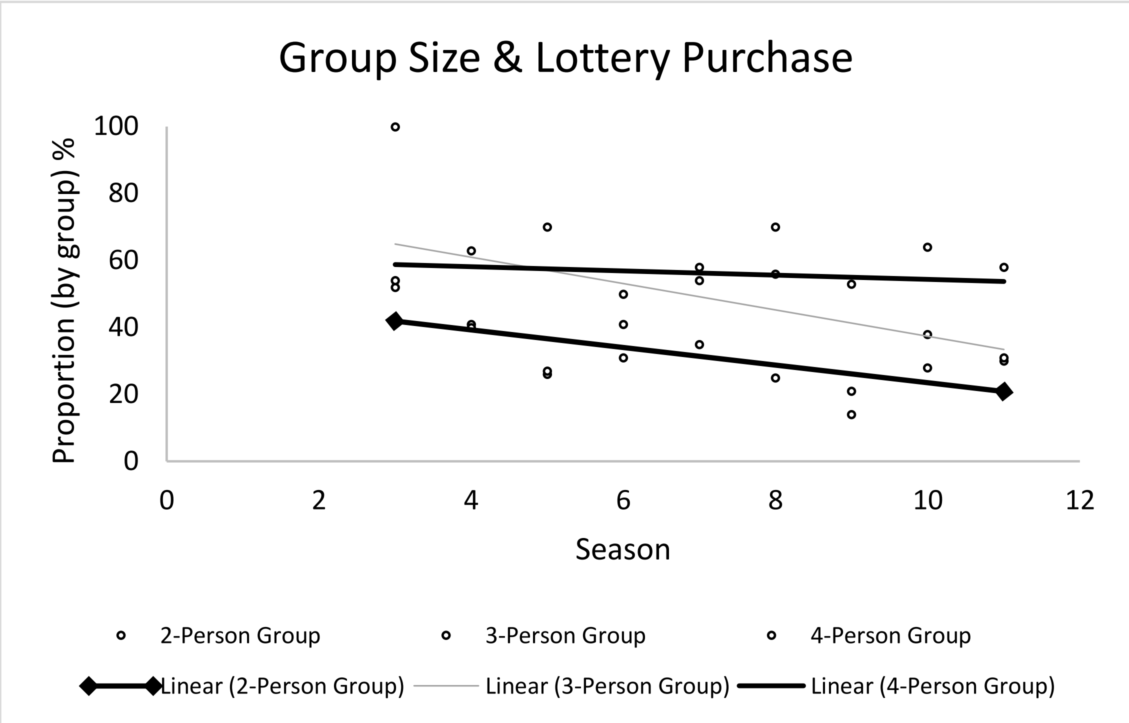


Figure 1.3
Group size effect on lottery purchase at the RIDP
This figure plots the proportions of the participant groups in our sample that purchase the lottery upon arrival at the RIDP but, in this case, with the trend lines activated by group size. The RIDP concept is described in the main text. The proportions for the respective group sizes are thus based on 63.1% of the total 1047 groups in our main sample. See Table 1.1 for more details on the summary statistics. See Appendix 1.1 Table 1 for an outline of our main variables and the unique source of data. The trend lines indicate not only time effects but also that, on average, more 3 or 4-person groups arriving at the RIDP purchase the lottery than do the 2-person groups. The latter differences appear somewhat consistent across seasons. This pattern of increasing tendency to purchase the lottery with group size, though slightly diminished, is also observed between the 3 and 4-person groups, except for a minor reversal before season 5. The shy reversal there may well be due to other factors relevant to the decision to purchase the lottery.

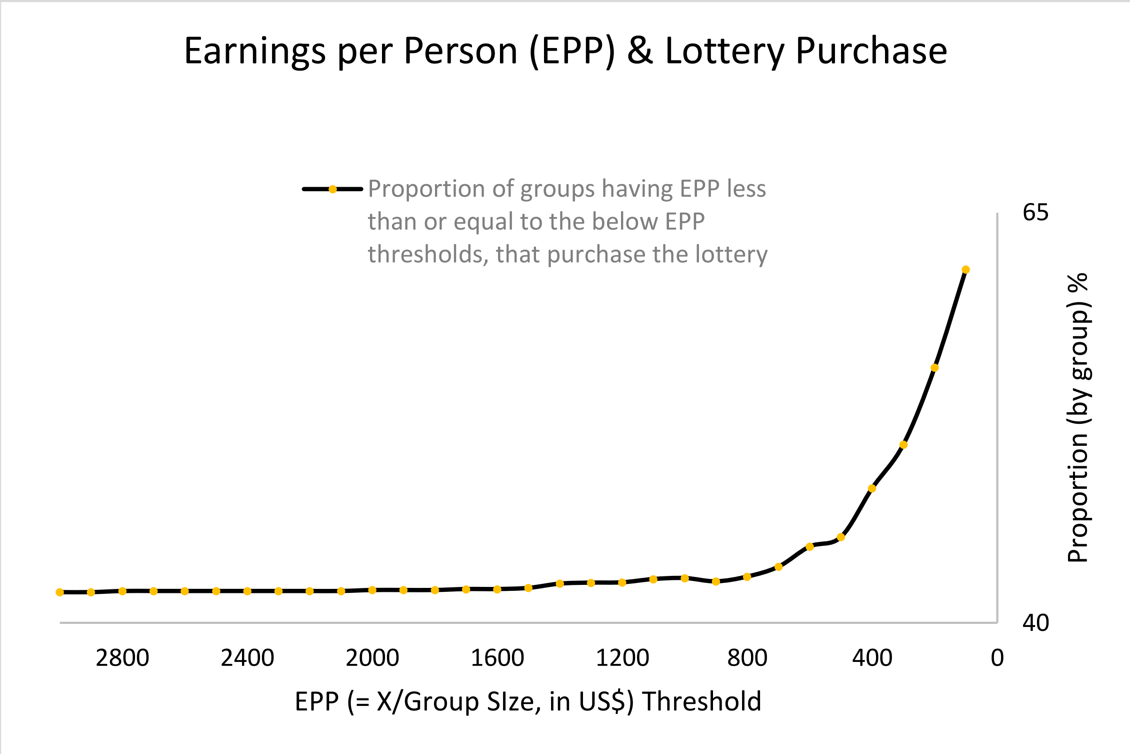


Figure 1.4
Earnings per person (EPP) effect on lottery purchase at the RIDP
This figure plots the proportions of the participant groups in our sample (having earnings per person, EPPs, equal to or below a threshold) that purchase the lottery upon arrival at the RIDP. The RIDP concept is described in the main text. The proportions are based on 63.1% of the total 1047 groups in our main sample. See Table 1.1 for more details on the summary statistics. See Appendix 1.1 Table 1 for an outline of our main variables and our unique source of data. For each threshold, we collect the set of all groups with EPP less than or equal to that threshold and then compute the proportion of groups in that subset that purchase the lottery upon arrival at the RIDP. The flat line portion of the graph mimics the frequency distribution of the EPP but indicates the average lottery purchase rate, conditional on arrival at the RIDP is about 42% for the entire sample – a trend we show in Figure 1.2 to be decreasing across time. More importantly, however, the graph also reveals some earnings and group size effects that appear to exacerbate below an EPP of $400.

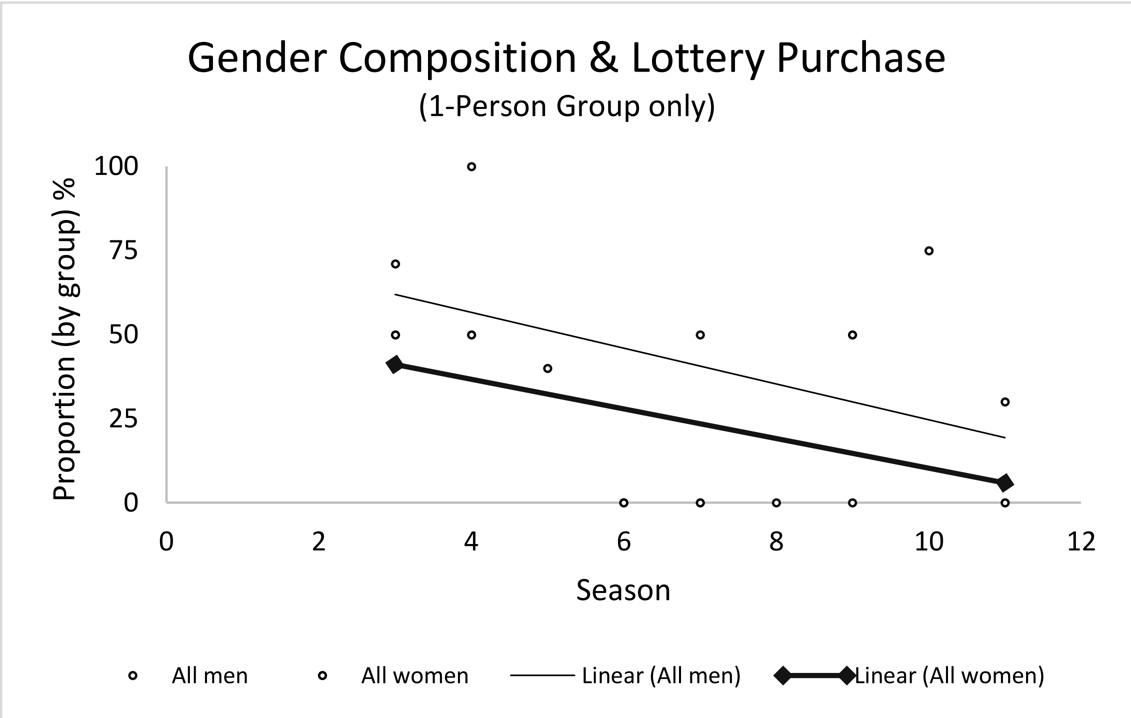


Figure 1.5
Gender composition effect on lottery purchase at the RIDP for 1-person groups
This figure plots the proportions of the participant groups in our sample that purchase the lottery upon arrival at the RIDP. The proportions are based on the 1-person groups in our main sample. See Table 1.1 for more details on the summary statistics. The RIDP concept is described in the main text. See Appendix 1.1 Table 1 for an outline of our main variables and our unique source of data. We compute the proportions for only 1-person groups. The trend lines indicate the earlier discussed time effect and, with group size fixed, a near-perfect time-consistent difference between a 1-male committee and a 1-female committee. Individual males appear to want to purchase the lottery more often than do individual females.

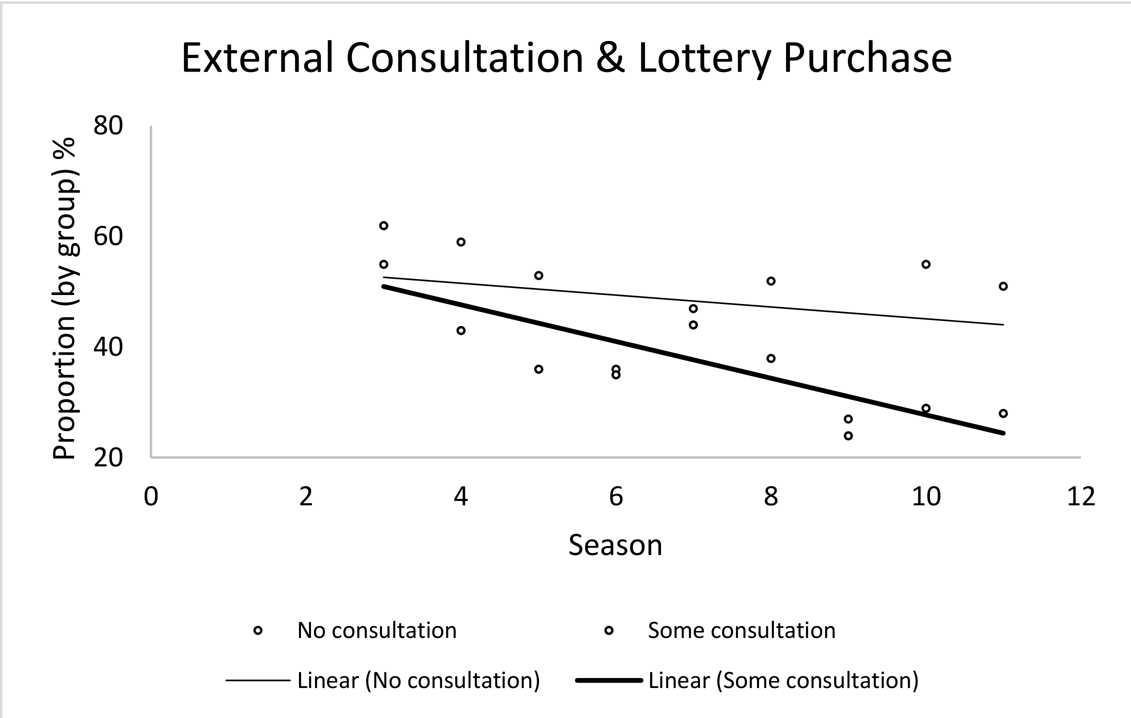


Figure 1.6
External consultation effect on lottery purchase at the RIDP
This figure plots the proportions of the participant groups in our sample that purchase the lottery upon arrival at the RIDP but with the data split in two: A set comprising of all groups arriving at the RIDP that did not use any external consultation and the other comprising of all groups arriving at the RIDP that did use some external consultation. The RIDP concept is described in the main text. The proportions are thus based on 63.1% of the total 1047 groups in our main sample. See Table 1.1 for more details on the summary statistics. See Appendix 1.1 Table 1 for an outline of our main variables and our unique source of data. The trend lines appear to indicate that the groups that did use some external consultation tend to purchase the lottery less often than those that did not use any external consultation. The differences appear to widen across seasons.

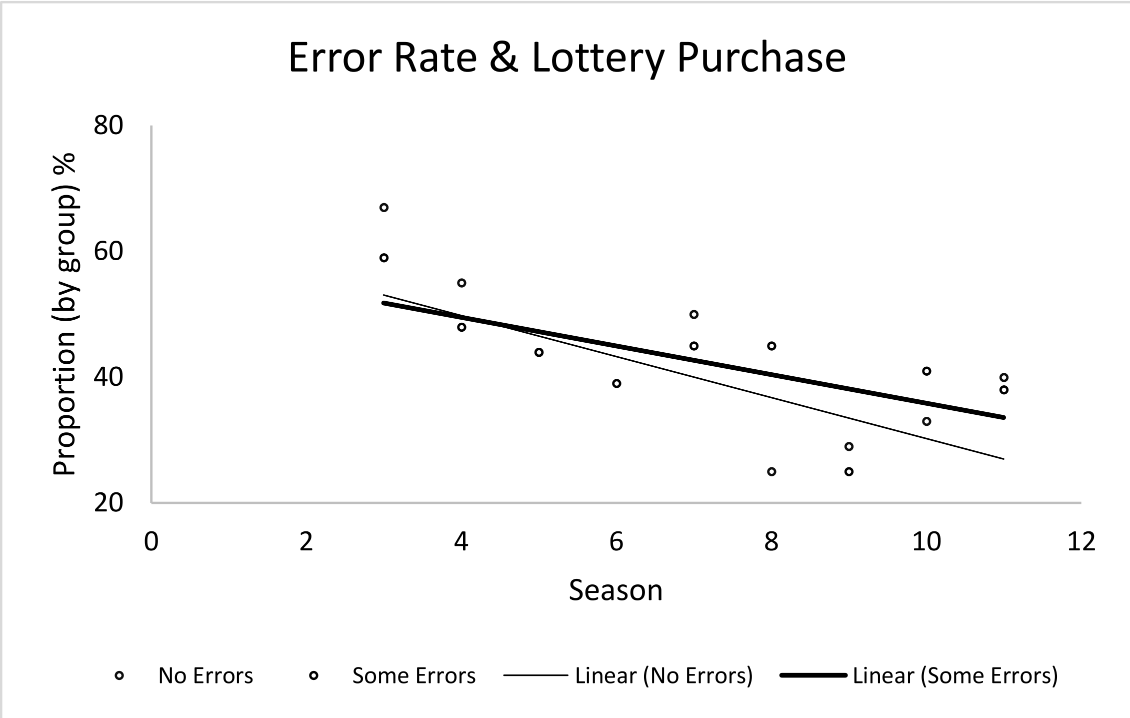


Figure 1.7
Error rate effect on lottery purchase at the RIDP
This figure plots the proportions of the participant groups in our sample that purchase the lottery upon arrival at the RIDP but with the data split in two: A set comprising of all groups arriving at the RIDP that had zero errors and the other comprising of all groups arriving at the RIDP that had some errors. The RIDP concept is described in the main text. The proportions are thus based on 63.1% of the total 1047 groups in our main sample. See Table 1.1 for more details on the summary statistics. See Appendix 1.1 Table 1 for an outline of our main variables and our unique source of data. The trend lines appear to indicate that the groups that had some errors tend to purchase the lottery more often than those that had zero errors. The differences appear to widen with time, but appear much smaller compared to the external consultation effect.

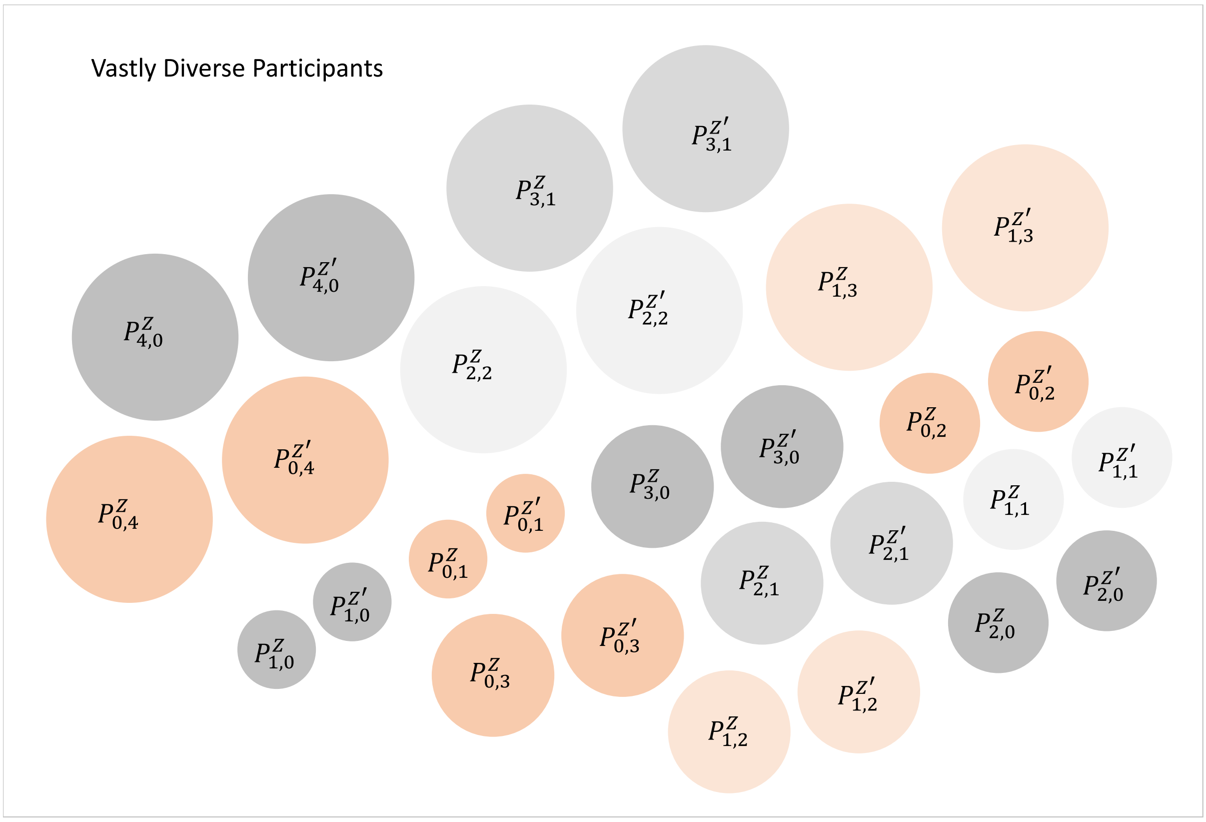


Figure 1.8
Diverse *Cash Cab* game participants and consolidation of pure gender effects
This figure highlights that the *Cash Cab* participants are vastly diverse in demographic characteristics such as age, gender, race, status, etc. Unassuming participants, $P_{m,f}$, with m number of males and f number of females board a cab to head to a destination and ultimately learn that they are in a game show. If they agree to play, they work to answer questions and earn a cumulative sum, X. The symbols Z and Z' represent opposite characteristics (e.g., young v. old; white v. black; and many more). Even though the choice to participate in the game is not a random assignment by me as the researcher, the deep diversity of the participants allows me to regroup them along progressive shades of gender categories—e.g., Homogenous Male, Majority Male, Gender-Balanced, Majority Female, and Homogenous Female. These pooled groups with similar gender categories but multiple opposite characteristics allow me to compute the average impact of gender diversity on risk-taking for each subgroup, by re-enforcing the pure gender effects while averaging out other subgroup characteristics. See Figure 1.9 for more details of the overall identification strategy.

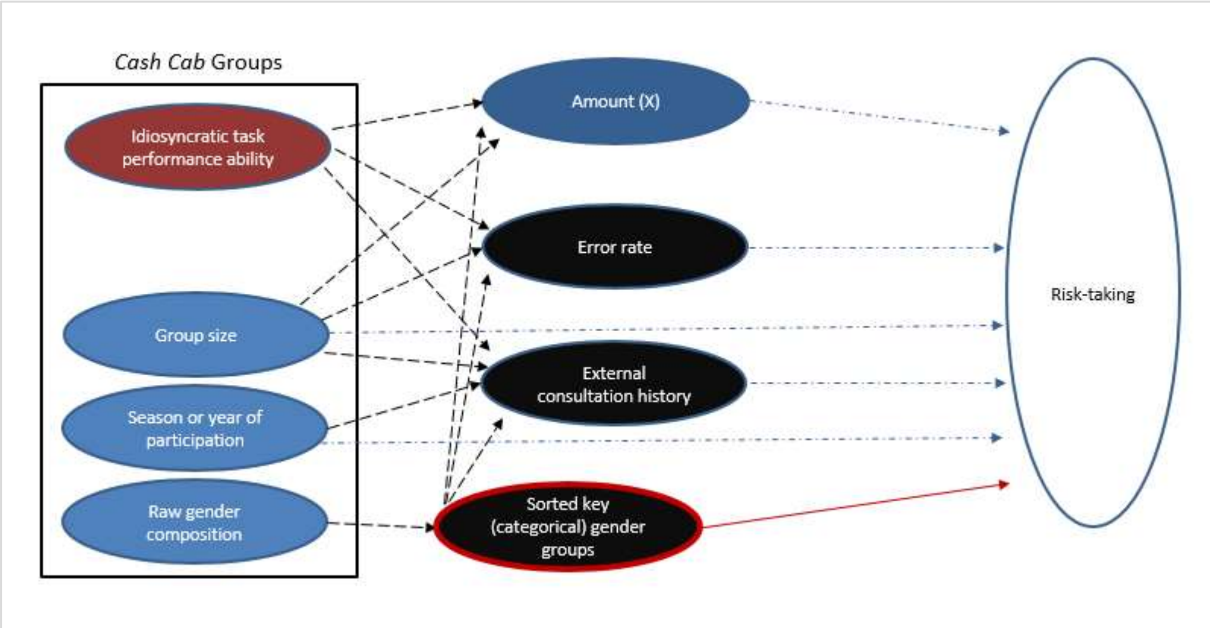


Figure 1.9
Gender composition and risk-taking in the *Cash Cab*: Identification strategy and assumptions
This figure shows the identification strategy and assumptions in the *Cash Cab* pseudo-natural laboratory setting: Each participating group in the *Cash Cab* has its idiosyncratic task performance ability (linearly or non-linearly derived from the individual task performance abilities of the group members); year of participation, group size; raw gender composition; and other random effects that might contribute to the absolute risk-taking behavior of the group. Moreover, the idiosyncratic task performance ability of each group is not directly observable. To alleviate much of the identification challenges, first, we sort the raw groups into key gender categories—homogeneous male, majority male, gender-balanced, majority female, and homogeneous female. The sorting of the groups into broad gender categories is aimed at not only focussing the general empirical analysis on the relative differences in risk-taking between the gender groups but also (by so doing) averaging out other potential random effects that may not have been explicitly measured. Next, because the unobservable idiosyncratic task performance ability of each group contributes directly to the group's error rate, the group's tendency to consult externally, and the amount won by the group, we construct binary variables from the latter observable characteristics to account for the former. The group size and the year of participation in the game are directly observable, hence we include them as controls to account for the fixed effects associated with the size of each group and the potential viewer learning effects that might have occurred across seasons since the game first aired on Cable TV in the USA. Consequently, our identification assumption is that after controlling for the game season, group size, error rate, external consultation history, amount won; and sorting the raw groups into the broad gender categories—homogeneous male, majority male, gender-balanced, majority female, and homogeneous female—any residual differences in the risk-taking behavior of the groups are, on average, caused by the differences in their broad gender group types.

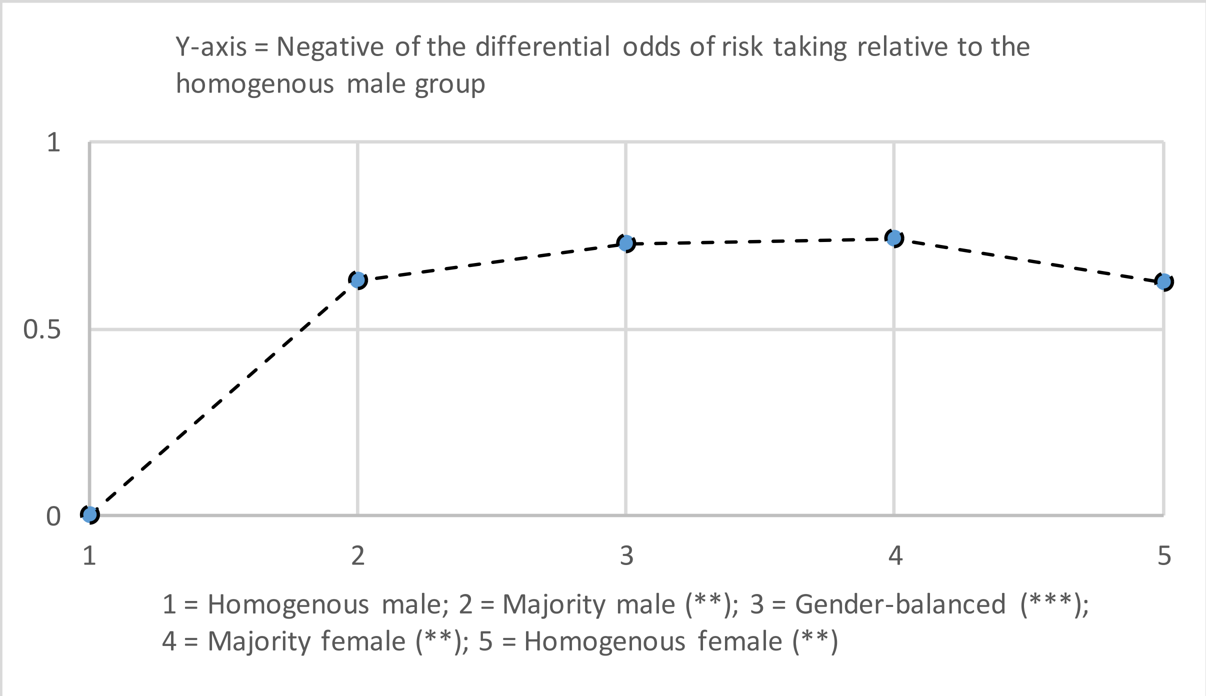


Figure 1.10
Difference in Log Odds of Risk-Taking (Before Hypotheses Tests)
This figure plots the differential regression coefficients (in Table 1.3, Model 1) for the gender subgroups with the homogeneous male group as the reference group. It would seem as if all gender diverse groups—be they majority male, gender-balanced, or majority female—are each potentially less likely than (or at least as likely as) a homogeneous female group to purchase the lottery at the RIDP. The latter observation suggests some non-linear (or most likely concave) gender diversification effect on the groups' collective risk aversion. Consequently, using Model 1 as the reference multiple logistic regression model, we conduct a series of tests to evaluate the following set of null hypotheses:

- With the all-male group as the reference group, the differential effects of each of the diverse subgroups are not individually and significantly different from that of an all-female group

Appendix 1.1 Table 2 reports the p-values for the hypotheses tests. In all, we fail to reject the respective null hypotheses since the reported p-values are all greater than the significance level of 5%.

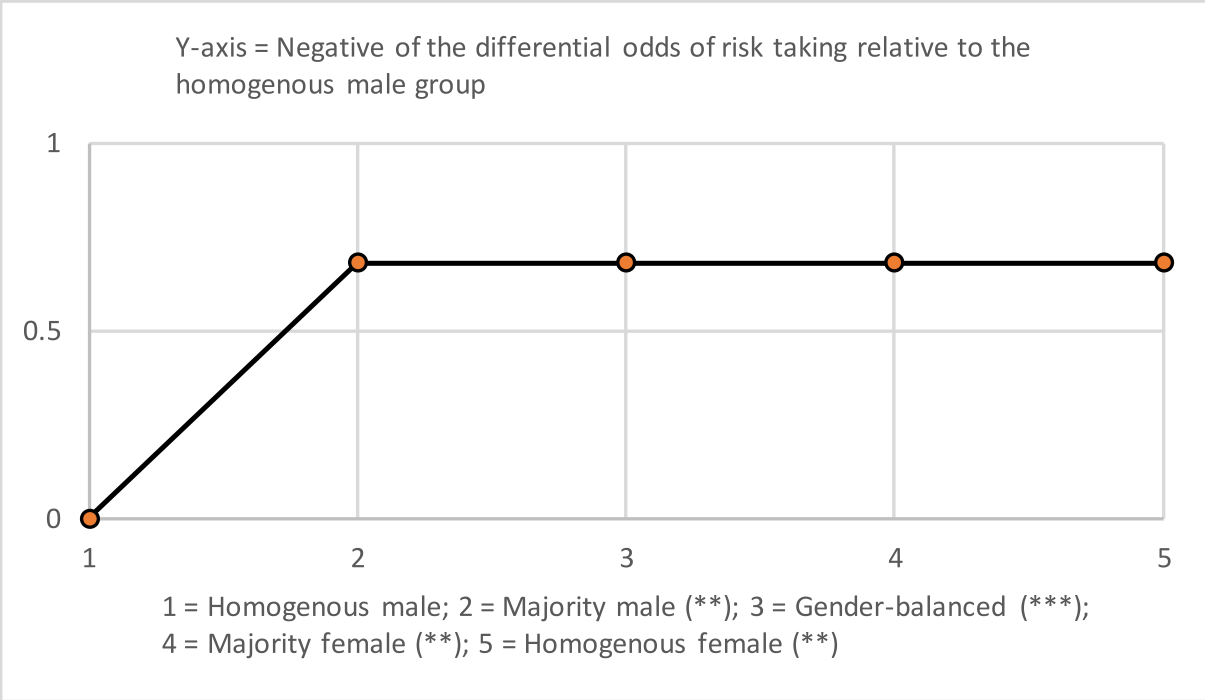


Figure 1.11
Difference in Log Odds of Risk-Taking (After Hypotheses Tests)

Following the results of the hypotheses tests in Appendix 1.1 Table 2, we update the plot of the differential regression coefficients of the gender subgroups previously shown in Figure 1.10. A basic look at this updated plot might lead to the interpretation that a homogeneous female group will be financially more risk-averse than a homogeneous male group, all other things equal. Also, a simple gender-diverse group—be it majority female, gender-balanced, or majority male—will produce a level of increased risk-aversion relative to that of a homogeneous male group that is at least as high as that which is likely to be produced by a homogeneous female group, all other things equal. It would appear, however, that locked within this picture of some acute concavity of the gender-diversification effect is a more fundamental and analytical interpretation of the results that relies hugely on the critical examination and full integration of the key features of the game setting (e.g., the individual veto power implicit in the game setting).

Table 1.1
Summary statistics (5 out of 6)
This table reports summary statistics of 5 of the 6 key variables for our main sample. Definitions of all variables are provided in Appendix 1.1 Table 1.

	Observation	Proportion (%)	Total number of persons
Group Size			
1-person group	127	12.1	127
2-person group	453	43.3	906
3-person group	283	27.0	849
4-person group	184	17.6	736
Total	1047	100	2618
Gender composition			
All male groups	294	28.1	NA
Majority male groups	142	13.6	NA
Gender balanced groups	291	27.8	NA
Majority female groups	112	10.7	NA
All female groups	208	19.9	NA
Total	1047	100	NA
External consultation history			
No consultation	382	36	NA
Some consulation	665	64	NA
Total	1047	100	NA
Error rate			
No error	96	9	NA
Some error	951	91	NA
Total	1047	100	NA
Decision			
Rejected offer	384	58	NA
Purchased lottery	277	42	NA
Total	661	100	NA

Table 1.2
Summary statistics (1 out of 6)

This table reports summary statistics of the remaining 1 of the 6 key variables for our main sample: The Earnings variable. Approximately 36.9% of all groups in our sample had errors greater than the maximum number of errors allowed per group before reaching their respective destinations and, as a result, exited the *Cash Cab* forfeiting all of their cumulative earnings. Consequently, only about 63.1% of all groups arrived at the RIDP (i.e., got to the end of the game to get the chance to either purchase the risky lottery – bet their entire cumulative earnings in a double or lose all proposition – or happily part with their earnings intact). Again, definitions of all variables are provided in Appendix 1.1 Table 1.

Group Earnings (X, in US$)	Observation	Proportion (%)
[0, 0]	386	36.9
[1, 400]	80	7.6
[401, 800]	279	26.6
[801, 1200]	190	18.1
[1201, 1600]	73	7.0
[1601, 2000]	20	1.9
[2001, 2400]	12	1.1
[2401, 2800]	4	0.4
[2801, 3200]	3	0.3
Total	1047	100

Note: Approximately 63.1% of the groups in the sample arrived at the "risky table"

Table 1.3
Effect of gender diversity

This table reports the standard multiple logistic regression of the decision variable on the gender composition variable. The specific intuitive equation is given in the main text. I, however, control for other important group characteristics such as the season (or year), group size, the group cumulative earnings (X), and the external consultation history. We report two sets of results: Model 1 without, and Model 2 with, the groups' error rates. We ultimately disregard Model 2 in future analyses since the error rate is not significant at 1%, 5% or 10% (not shown) in our main sample and our key results remain unchanged in both models.

Dependent variable = Log odds of Decision		
	Model 1	Model 2
Homogeneous male group	0.0000	0.0000
	(0.0000)	(0.0000)
Majority male group	-0.6299	-0.6244
	(0.3075)**	(0.3079)**
Gender-balanced group	-0.7288	-0.7417
	(0.2573)***	(0.2580)***
Majority female group	-0.7420	-0.7375
	(0.3398)**	(0.3401)**
Homogeneous female group	-0.6262	-0.6207
	(0.2981)**	(0.2984)**
Controls:		
(1) Season (year)	Yes	Yes
(2) Group size	Yes	Yes
(3) Cumulative earnings	Yes	Yes
(4) External consultation history	Yes	Yes
(5) Error rate	No	Yes
_cons	1.3153	1.1473
	(0.3972)***	(0.4685)**
Pseudo R-Squared	0.094	0.095
N	599	599

* $p<0.1$; ** $p<0.05$; *** $p<0.01$
Standard errors in parentheses

Table 1.4

Group size fixed effects—Does Group Size Always Matter for Risk-Taking?

This table reports the standard multiple logistic regression of the decision variable on the group size variable. The specific intuitive equation is given in the main text. I, however, control for other important group characteristics such as the season (or year), gender composition, the group cumulative earnings (X), and the external consultation history. We report two sets of results: Model 1 without, and Model 2 with, the groups' error rates. We ultimately disregard Model 2 in any further analyses since the error rate is not significant at 1%, 5% or 10% (not shown) in our main sample and our key results remain unchanged in both models. We also report below a test of the null hypothesis that there is no difference in risk-taking between a 3-person group and a 4-person group, ceteris paribus, at the significance level of 5%.

Dependent variable = Log odds of Decision		
	Model 1	Model 2
2-person group	0.0000	0.0000
	(0.0000)	(0.0000)
3-person group	1.0501	1.0384
	(0.2680)***	(0.2687)***
4-person group	1.3610	1.3575
	(0.2697)***	(0.2698)***
<u>Controls:</u>		
(1) Season (year)	Yes	Yes
(2) Gender composition	Yes	Yes
(3) Cumulative earnings	Yes	Yes
(4) Overconfidence proxy	Yes	Yes
(5) Bounded rationality proxy	No	Yes
_cons	1.3153	1.1473
	(0.3972)***	(0.4685)**
Pseudo R-Squared	0.094	0.095
N	599	599

* p<0.1; ** p<0.05; *** p<0.01
Standard errors in parentheses

<u>Hypotheses:</u>

H_0: 3-person group = 4-person group

H_A: 3-person group ≠ 4-person group

<u>Result:</u>

(1) 3-person group — 4-person group = 0

chi2(1) = 1.42

Prob > chi2 = 0.2330

<u>Conclusion:</u>

Since Prob > chi2 = 0.2330, we fail to reject the null hypothesis that, on average, there is no difference in risk-taking at the RIDP between a 3-person group and a 4-person group, ceteris paribus, at the significance level of 5%.

Appendix 1.1 Table 1
Variable definitions: *Cash Cab*

Variable Name	Source	Definition
Group Size	*	Categorical variable equal to n (=1, 2, 3 or 4) if the number of persons in a group is n
Gender composition	*	Categorical variable equal to 1, 2, 3, 4, 5 for All Male, Majority Male, Gender Balanced, Majority Female and All Female groups, respectively
External consultation history	*	Indicator equal to 0 if a group did not use its "consultation budget" and 1 otherwise
Error rate	*	Indicator equal to 0 if a group arrived at the RIDP with zero error and 1 otherwise
Earnings	*	Cumulative earnings (X) of the group in US dollars
Season	*	Time (year) categorical variable for the 9 seasons
Decision	*	Indicator equal to 1 if a group accepted the offer at the RIDP and 0 otherwise

* *Cash Cab*, seasons 3 to 11.
https://www.discovery.com/tv-shows/cash-cab/

Appendix 1.1 Table 2
Hypotheses tests for Model 1

This Table reports the p-values for the tests of hypotheses of the gender composition categories (on a row vs column basis) using Model 1 as the reference model. The All-Male column simply reproduces the actual p-values shown in the model, all of which were significant. We fail, however, to reject the hypotheses that, compared to an all-male group, the differential effects of the diverse subgroups are not individually and significantly different from that of an all-female group, since the p-values of those tests are respectively greater than the significance level of 5%. Overall, these results seem to imply that compared to a homogeneous male group, any diverse subgroup is at least as likely as an all-female group to take fewer risks when faced with a setting similar to the conceptual RIDP, all other things equal.

		Prob > Chi2				
S/No.	Gender Composition	All Male	Majority Male	Gender Balanced	Majority Female	All Female
1	All Male					
2	Majority Male	0.04				
3	Gender Balanced	0.01	0.76			
4	Majority Female	0.03	0.72	0.97		
5	All Female	0.04	0.99	0.73	0.75	

Significance level = 0.05

Appendix 1.1 Table 3
References to easily accessible *Cash Cab* video data

S/N	How to access this video data	Description	Relevance	Remark
1	GSN (online), Season 3, Episode 25 **<u>Caption</u>** "Men with women **not** risk-averse"	A 3-person group consisting of 2 males and 1 female. The woman in the group was the first to declare her risk-aversion but both men wanted to take the risk and gently persuaded her to join them. She did not insistently say no, hence they took the bet collectively. Their EPP was $483.33	This game is one of many games that show that the conjecture that <u>*men in non-homogenous male groups may be at least as risk-averse as the women in their group*</u> is not an important feature of the *Cash Cab* setting.	Their EPP of $483.33 is greater than the median EPP of the entire data in our sample.
2	GSN (online), Season 13, Episode 10 **<u>Caption</u>** "Male veto power/influence in action"	A 2-person group consisting of 2 males (a father, possibly in his late 50s to mid-60s and his early teenage son). The father was decidedly risk-loving but the son vetoed.	This game is one of many games that show the importance of the individual veto power implicit in the game setting.	Their EPP of $575 is greater than the median EPP of the entire data in our sample
3	YouTube, *Cash Cab* 3.12.12 **<u>Caption</u>** "Female veto power/influence in action"	A 3-person group consisting of 2 males and 1 female. The woman was the first to declare her risk-aversion and insistently said no to the bet. She also actively sought to influence the group with some generally wrong statistics about final bets in the *Cash Cab*.	This game is one of many games that show the importance of the individual veto power implicit in the game setting.	Their EPP of $150 is less than the median EPP of the entire data in our sample

Chapter 2

How Many Female Seats on a Board? Board Gender-Diversification, Power, Risk-Taking, and Financial Performance

2.1. Introduction

Board gender-diversification (i.e., the systematic increase in the proportion of board seats occupied by female directors) remains an important subject across the globe.[18] In the USA, support for more female directors on boards is no longer a mere matter of principle or charity.[19] Yet, the economic, the business, or simply, the non-social-justice-and-equity-based case for gender diversity on corporate boards remains inconclusive (e.g., see Ahern and Dittmar, 2012; Matsa and Miller, 2013; Rhode and Packel, 2014; Eckbo et al., 2019). While practitioners continue to have mostly positive views[20] of the impact of gender diversity on corporate behavior and firms' financial performance, the empirical evidence in the academic literature continues to be mixed, with the associated discussions and debates seemingly intensifying rather than abating.[21] Buoyed by the mixed extant

[18] Notably, in 2006, Norway—a non-European Union (EU) member—pioneered the imposition of a 40% minimum female gender quota on the corporate boards in its jurisdiction. By 2014, at least sixteen countries had followed suit and, in 2017, the EU began pushing for similar quotas across the entire region.
[19] In 2017, the "Big Three" institutional investors push for greater board gender diversity (e.g., see Gormley, et al, 2020). In 2018, California became the first U.S. state to pass legislation requiring publicly traded firms to have at least one woman on the board by 2019, with more mandates to increase the quota in line with the size of the board by 2021 (e.g., see Franceschet and Piscopo, 2013; Zillman, 2017; Ang, 2019). Also "in January 2020, the Nasdaq Stock Market filed a proposal with the Securities and Exchange Commission to adopt Rule 5605(f) (Diverse Board Representation), which would require Nasdaq-listed companies, subject to certain exceptions, to have (A) at least one director who self-identifies as a female, …" (Rau, et al., 2021).
[20] "Goldman Sachs CEO says it won't take a company public without diversity on its board" – McGregor (2020), Washington Post, January 23.
[21] "Why Board Diversity and the Nasdaq Rule Requiring It Make Sense" – Painter (2021), Columbia Law School's Blog on Corporations and the Capital Markets, April 26.

academic evidence, these debates, frequently spirited, also permeate the general public space, with significant implications for government policies and legislation.[22]

This chapter provides a novel set of empirical analyses and results to shed new light on the debates. We argue that the diverse nature of the extant empirical evidence in the academic literature is strongly associated with the difficulty in empirically isolating how different board structures affect the impact of board gender diversity on firm risk-taking and performance. We focus on the differential roles of individual power and influence asymmetries within-board. Our empirical methodology thus hinges on concerted attempts at isolating and untying these differential roles of power/influence in the corporate board setting. In sum, we conjecture[23] as follows: if a female director is unlikely to have any personal power or influence on the board, her addition to the board will have no significant impact on firm risk-taking and financial performance. However, with increasing power and influence on the board via (a) greater numerical strength and/or (b) non-token aggregate position, female directors will tend to reduce the excessive risk-taking behavior of the firm and, to the extent that the gender-diversification is non-disruptive, this effect can feature significant increases in profitability and firm value. Indeed, these hypotheses appear pretty intuitive and easy to think through.

Ex-ante, however, our view is that if the subsequent empirical tests show consistent results, then the hypotheses would be largely reconciliatory. We hold this somewhat optimistic view for several reasons. One reason is that the hypotheses are consistent with the findings of Wabara (2021a), who uses a unique (pseudo-laboratory) context to show

[22] "We commend individual firms for the proactive efforts they have already made in recruiting, promoting, and maintaining diverse talent," the senators, led by Pennsylvania's Pat Toomey, wrote. "However, it is not the role of Nasdaq…to act as an arbitrator of social policy or force a prescriptive one-size-fits-all solution upon markets and investors." – Ackerman (2021), Wallstreet Journal, Feb. 12, 2021. Despite strong internal and external opposition, the SEC ultimately approved the Nasdaq rule on August 6, 2021. For details on the latter, see https://www.sec.gov/rules/sro/nasdaq/2021/34-92590.pdf.

[23] See Section 2.2 for more details.

that, "on average, a female will take less *excessive*[24] financial risks (i.e., mainly when the ex-ante expected profit is zero). However, in small groups, an individual's (or a subgroup's) share of the within-group power and influence distribution, not just numerical strength, determines whether their tendencies would manifest in the collective decisions or outcomes of the entire group." A significant insight from Wabara (2021a) is that self- or sub-group power/influence guarantees voice and inclusion (and, therefore, impact within-group). Another reason is that the hypotheses' power/influence asymmetry aspect is also consistent with the primary assumption in Adams, Almeida, and Ferreira (2005) and the central theme of the minority relations literature (e.g., Gittler, 1956; Noel, 1968; Yetman, 1985). More so, the profitability and firm value components are consistent with the results of Berliant and Fujita (2012), who theoretically model diversity as endogenous and immobile but demonstrate how multiple cultures interacting can improve productivity.

Therefore, a significant contribution of this study consists of designing and using a novel combination of empirical tools and analyses to test our largely intuitive set of hypotheses using the conventional board and financial data. In effect, to achieve this goal, we adopt a quasi-experimental strategy involving automated searches for endogenously structured treated and control samples. However, we deal strategically with the threat of endogeneity problems such as self-selection issues and ex-ante board beliefs and motives for director selections (e.g., see Hermalin and Weisbach, 1988; Agrawal and Knoeber, 2001; Bertrand and Schoar, 2003). Precisely, we measure the average differential impact of gender by computing stacked difference-in-differences estimates on multiple sets of

[24] We highlight, in Section 2.2, the reconciliatory nature of the Wabara (2021a) results concerning the relationship between gender and individual risk preferences. Specifically, while every female need not be less risk-loving than any male, the results suggest a crucial condition (i.e., when the risk-taking is qualitatively excessive) under which female directors might be more risk-averse in a value-maximizing way for the firm.

(structured quasi-temporal event counterfactual) board and financial data samples for publicly listed US firms. We also progressively structure our counterfactual sample searches so that our results, if sequentially consistent, would severely weaken pure endogenous selection stories and strongly enhance our power and influence hypotheses.

We are primarily concerned about the issue of endogenous selection based on the compatibility of risk preferences, as it can lead to reverse causality. Endogenous selection based on the compatibility of risk preferences may mean that the female directors are chosen, not because of their ability to influence board decisions and firm policy toward reducing excessive risk, but because of the ex-ante compatibility of their risk preferences with those of the incumbents. We specifically work to rule out such reverse causality concerns. We are less concerned about the issue of endogenous selection based on ability (e.g., financial expertise, etc.) because it is not necessarily incompatible with our power and influence hypotheses. For instance, suppose female directors are chosen for their ability to reduce excessive financial risk. In that case, such an ex-ante motive will bolster our hypotheses, more so if we find consistent empirical evidence that the ability to reduce excessive financial risk mainly manifests with increasing power and influence.

To compute the stacked difference-in-differences estimates on the quasi-temporal event counterfactual samples, we adapt the main functional form in Bertrand and Mullainathan (2003) by using a fixed window and saturating the model with firm and industry-year fixed effects. We also evaluate both firm-level policies and long-term market reactions. As proxies for the firm-level determinants of risk, we use firm leverage, operating profitability, cash-asset ratio and its volatility, and tangibility.[25] For the market-level risk factor, we use the volatility of firms' market-adjusted returns. Also, following

[25] We use these proxies because most of them have well-documented stylized relationships in the capital structure literature (e.g., see Rajan and Zingales, 1995) and will help sanity-check the consistency of our results.

Adams and Ferreira (2009), Faleye, Hoitash, and Hoitash (2011), to mention just two, we use Tobin's q to proxy for firm performance.[26] We further outline in the discussion section how we rule out such endogeneity concerns as board choice of female directors with ex-ante motives or female director self-selection to the more aggressive excessive risk-reducing firms. Ultimately, we present our results as answers to crucial questions.

The first question we address empirically using our methodology is, does tokenistic board gender diversity affect firm risk-taking and performance? Our ex-ante expectations derive from the first part of our hypotheses. Hence, we do not expect any statistically significant changes in the firm-level risk proxies. We also do not expect any statistically significant changes in the firm value proxy or the volatility of the market-adjusted return. More so, for the latter because adding just one ordinary female director to the board would not have created much effective market communication during the period that our sample covers.[27] As we expected, we observe statistically non-significant changes across the board for all of the indicators, providing support for, at least, the first part of our hypotheses.

The next question that we address is, what happens when the numerical strength of the female subgroup of directors on the board is increased? Again, our ex-ante expectations derive from the second part of our hypotheses. Accordingly, in this case, we expect statistically significant changes in the firm-level risk proxies. Moreover, as Leary and Roberts (2005) point out, the dynamic pecking order theory predicts that firms are more concerned about excessively high leverage than excessively low leverage. That prediction is also consistent with Wabara (2021a). Hence, to the extent that a firm's existing leverage is excessive, we would expect female directors to use their increasing

[26] However, given the alternative definitions of the proxy (e.g., as a measure of investment opportunity) and the documented limitations of Tobin's q as a measure of firm value (e.g., see Bartlett and Partnoy, 2020), we also look at Peters and Taylors (2017)'s Total q. Our results remain qualitatively unchanged.
[27] Base Sample 1, used for our first two tests, covers the fiscal years from 2006 to 2019. Base Sample 2, used for our last two tests, covers the fiscal years from 2003 to 2019. However, much of the samples occur before 2016.

influence on the board to call for a reduction of the excess. We would also expect the changes in the other firm-level risk indicators to be consistent with Rajan and Zingales (1995). For example, if leverage is reduced significantly, we might expect a decrease in tangibility and increases in operating profitability and Tobin's q (our proxy for firm value). Also, if leverage is reduced significantly, we would normally expect a reduction in the volatility of the market-adjusted return; however, given the period that our base sample covers, this reduction may not be statistically significant. Again, as expected, we find results consistent with our conjectures. Interestingly, we also find that the directional changes in the proxies are consistent with Rajan and Zingales (1995).

While the latter gives us greater confidence in our stacked regression estimator, the question that naturally follows is the following: since our results are consistent with Rajan and Zingales (1995), might they also coincide with some equity market timing? This question is vital because Baker and Wurgler (2002) suggest an "equity market timing" theory of capital structure where the relationship between leverage and Tobin's q flows through net equity issues. The threat relates to how we might then interpret our gender effects. Suppose the effects we find are also consistent with Baker and Wurgler (2002). In that case, we might have a situation where the addition of female directors to the board could have been driven by the implications for the success of the fundraising activities.

The latter, although curious, may not be entirely farfetched, especially since Rau, Sandvik, and Vermaelen (2021) find that gender diversity boosts under-pricing in IPOs due to higher institutional investor demand for diverse boards. However, for this to matter in our case, our previous results should also be accompanied by an increase in net equity issue and not necessarily an increase in retained earnings. We document the contrary. Instead of a statistically significant increase, we find a statistically non-significant reduction in net equity issues. We also see a statistically non-significant reduction in net debt issues and a statistically non-significant change in total asset growth. However, we find an increase in newly retained earnings statistically significant at the 5% level.

Hence, our results are consistent with Rajan and Zingales (1995) but not compatible with Baker and Wurgler (2002), suggesting that the impacts of the increasing power and influence of the female directors (via greater numerical strength) on leverage and its determinants are not coincident with some equity market timing events.

The next question then is, through what channel are the female directors potentially influencing the managers to reduce excessive leverage? Stonehill (1975) documents that a goal of financial executives in many countries appears to be to guarantee the financial stability of their company and the availability of funds. This sentiment might explain the increased retained earnings, especially if the (now) more influential female directors emphasize such concerns to the managers. Strebulaev (2007) also suggests that firms might sell assets to pay down debt to help nip future liquidity crises in the bud. If the latter is the case, then it may be that the assets sold are the less productive ones since firm value and profitability improve, and we do not observe a statistically significant change in total asset growth. Consistent with these conjectures, we find an increase in physical asset sales significant at the 5% level and an increase in asset turnover that is significant at the 10% level. In general, these results suggest an activated managerial action to minimize an excessive risk of, say a future, liquidity crisis and to improve profitability and firm value.

Next, we switch our focus to the non-token aggregate position of the female subgroup of directors on the board and ask, what happens when the aggregate power of the female directors is increased? Here we use CEO transitions to drive the tests. Typically, CEO transitions generate a substantial amount of news and market communication. Hence, we expect any changes to the market-adjusted return volatility due to a male-to-female CEO transition to be statistically significant. In this case, an effective test would take care to evade any confounding effects from increased power/influence for the female directors via greater numerical strength, documented above. We adjust our testing strategy accordingly and find results (including a net increase in tangibility and asset turnover) consistent with the notion of reducing excessive risk (e.g., see Aggarwal et al., 2010).

Next, we check whether our results change (qualitatively) when we include firms in the utilities and financial industry excluded from our analyses up to this point. In the corporate governance literature, it is standard practice to exclude firms in those industries because of significant differences in regulatory oversight that can limit the board's influence on firm-level policies. However, the functional form of our stacked DiD regression estimator is intentionally saturated with several fixed effects, including interactive industry-year fixed effects (to help deal with layers of heterogeneities across multiple dimensions, e.g., year, cohort, firm, industry by year). If the fixed effects do much of what we intend for them to do, then relaxing the convention above and some other requirement we imposed on the structured samples search process should not change our results, qualitatively at least. As expected, our results remain qualitatively unchanged, strengthening the support for our hypotheses and increasing our confidence in the test implementation methodology. The sum of our results bolsters the idea that female directors do not simplistically reduce risk. Instead, they minimize excessive risk.

Finally, we conduct some real ex-post analyses of whether our results are causal. We begin by questioning whether the statistically significant results we find when female directors acquire greater numerical strength (or more significant positional status) on the board are mainly driven, not by the increased power and influence, as hypothesized, but by an endogenous selection of the female directors based on the compatibility of their risk preferences with those of the incumbents. We also question whether the same results might have been driven solely by an endogenous ex-ante motive of selecting the female directors based on their ability to reduce excessive risk. In section 2.6.2, we show that the former (i.e., endogenous selection of the female directors based on the compatibility of their risk preferences with those of the incumbents) fails to explain our sequence of tests consistently. We also show that an endogenous ex-ante motive (alone) of selecting female directors based on their ability to reduce excessive risk is insufficient since the ability mainly manifests with increasing numerical strength or power status on the board. In sum, these results provide strong support for our hypotheses, and we find them to be causal.

This chapter makes several contributions. To the best of our knowledge, this study is the first to use our novel set of quasi-experimental empirical analyses (involving a sequence of structured quasi-temporal event counterfactual samples and the estimation of consistent multi-cohort stacked DiDs over fixed, short, symmetric event windows) to document not only that but also when and how board gender diversity can have positive impacts on firm performance and value. The extant empirical evidence on this subject has been mixed (e.g., see Zahra and Stanton, 1988; Erhardt, 2003; Adams and Ferreira, 2009; Wang and Clift, 2009; Ahern and Dittmar, 2012; Matsa and Miller, 2013; Gregory-Smith, 2014; Chen et al. 2019; Eckbo et al., 2019; et cetera). Our study provides new perspectives using novel empirical analyses. Principally, we shed new light on how the power/influence asymmetry within-board modulates the impact of gender diversity on firm performance. In this way, our study[28] helps explain (or at least contributes to the reconciliation of) the mixed extant empirical evidence in the academic literature.

[28] We also provide deeper insights into the functioning and robustness of our quasi-experimental empirical methodology. We do so using a simple theoretical model and numerous simulation results. We provide simple proofs and collect the simulation results in the Internet Appendix. Generally, our expositions draw inspiration from the intuition and language of the common midpoint (CMP) stacking methodology introduced by William Mayne in 1962 and used in seismic engineering and geophysical imaging to identify complex subsurface objects. Seismic engineering and geophysical imaging are two of this author’s previous jobs in the industry.

Overall, our results suggest that it might be more sustainable to begin moving the public (and the academic) discussions and debates from whether diversity, board gender diversity, in particular, adds value to how best might the gender-diversification processes occur to ensure that the value that is inherent in diversity can be profitably unlocked. Also, our results, while confirming vital stylized facts about the determinants of leverage (e.g., Rajan and Zingales, 1995), show that increased power and influence for the female directors (via greater numerical strength or non-token aggregate position) might help firms operate at optimal leverage levels since female directors do not just reduce risk; they reduce excessive risk and enhance firm value.

The rest of this chapter proceeds as follows. Section 2.2 leads with arguments that suggest that some negative impact of board gender diversity on corporate risk-taking and performance documented in the literature may have other compelling explanations. Section 2.3 details our model and empirical examinations. Section 2.4 presents the results. Section 2.5 presents additional analyses to evaluate the robustness of our methodology and findings. Section 2.6 discusses the implication of these findings. Section 2.7 concludes.

2.2. Does Gender Diversity Really Damage Firm Value?

In this section, we first present arguments that suggest that some negative impact of board gender diversity on corporate risk-taking and performance documented in the literature may have other compelling explanations, different from the actual presence of female directors on the board. Next, we argue that although a zero-impact might be closer to reality for US boards, there are conditions under which board gender diversity can affect corporate risk-taking and performance in a value-maximizing way.

2.2.1. Alternative explanation for some documented adverse effects of gender diversity

An inexhaustive list of some compelling alternative explanations (for the negative impact of board gender diversity on corporate risk-taking and performance documented in the literature) includes: (i) inadvertent triggering of board monitoring intensiveness, which by itself can reduce firm value (e.g., see Adams and Ferreira, 2009; Faleye, Hoitash, and Hoitash, 2011), (ii) the often disruptive nature of exogenous team assignments or exogenous gender diversification processes, such as gender quota laws (e.g., see Calder-Wang, Gompers, and Huang, 2021), and (iii) the possibility that the observed stock price reactions might not reflect investors' actual judgment on the presence of female directors on the board (e.g., von Meyerinck, et al., 2021). We, below, briefly discuss each of them.

2.2.1.1. An inadvertent triggering of board monitoring intensiveness

Using a sample of US firms and an instrumental variable methodology to deal with the potential threat of endogeneity, Adams and Ferreira (2009) summarize that female directors significantly impact board inputs and outcomes. Specifically, female directors have better attendance records than male directors, and male directors have fewer attendance problems the more gender-diverse the board is. They nevertheless document that the average effect of gender diversity on firm performance is negative. In other words, female directors show up more often to do their jobs on the board, and their greater dutifulness rubs off rather positively on their male peers. Still, the net impact (of their dutifulness and positive peer effect) reduces firm value.

To the authors' plaudits, Adams and Ferreira (2009) also acknowledged the difficulty with the interpretation and linked the result to a heightened monitoring intensiveness of the board when female directors join. However, in light of other evidence provided in the paper and new evidence (e.g., see Milbourn and Wabara, 2021), it is also easy to see why the negative effect on firm value may not directly result from the presence of female directors. For instance, Adams and Ferreira (2009) also show that women hold few corporate board seats and are more likely to join mainly two principal monitoring committees (i.e., audit and nominating/governance committees, but not the compensation committee). Hermalin and Weisbach (1988) and Agrawal and Knoeber (2001) document that firms appear to choose directors for their characteristics. The current debates to have more women on the board (and in more significant positions) suggest that the female directors have tended to have relatively minor aggregate power[29] on the board and may not have unilaterally self-assigned to the monitoring committees.

Nevertheless, once assigned to the monitoring committees, the evidence shows that female directors show up for work more dutifully and are thus more likely to straddle two principal monitoring committees. They also influence their male peers to show up more frequently to do their jobs. Coincidentally, even if only mechanically, this dutifulness enhances the likelihood that the board becomes monitoring-intensive. Indeed, intense board oversight has key benefits; but, its net impact typically reduces firm value (e.g., Almazan and Suarez, 2003; Adams and Ferreira, 2007; Faleye, Hoitash, and Hoitash, 2011). A crucial question then is, can a subgroup be to blame for dutifully doing a job they were assigned to do? Suppose the board's leadership assigns the (relatively few) female directors to other committees where their apparent dutifulness and peer effects would be less likely to activate the monitoring-intensiveness of the board. In that case, would the average impact of their presence on the board still lead to a loss in firm value? Suppose also that a homogenous male board was monitoring-intensive. Would the intense

[29] Adams and Ferreira (2009) document that only 25% of the firms in their data had more than one female director in 2003. We show, in Figure 2.1, that the average annual proportion of female directors on the boards of publicly listed US firms did not exceed 10% until after the financial crisis in 2008. Since then, although this annual proportion as a percentage of the board size has risen steadily, it appears to be peaking at around 22%.

oversight of a homogenous male board not also lead to a net negative impact on firm value? Moreover, Milbourn and Wabara (2021) show that the exact outcomes of intense oversight could differ depending on the relative sense of higher purpose of those leading the monitoring-intensiveness on the board.

2.2.1.2. The often disruptive nature of exogenous gender diversification processes

If exogenous shocks to board gender composition (such as gender quota laws) could trigger the required exogenous variations without other encumbrances, it would suffice to measure the differences in some appropriate risk-taking and performance proxies before and after the date of the enactment of the law. However, for studies on board gender diversity, this would be a problematic empirical proposition and may lead to mixed results, given several behavioral factors at play when the board re-composition is forced.

Take Norway's 2006 imposition of a 40% female gender quota on corporate boards as a case in point: Matsa and Miller (2013) show that the firms affected by the quota undertook fewer workforce reductions, had higher labor costs and employment levels, and had lower short-term profits than comparison firms. Ahern and Dittmar (2012) conclude that the constraint imposed by the quota led to a significant drop in the stock prices of the affected firms at the announcement of the law; a large decline in Tobin's Q over the following years; younger and less experienced boards; increases in leverage and acquisitions; and deteriorations in operating performance. However, Eckbo et al. (2019) disagree, arguing that the gender quota had no net negative effect on the affected firms and suggested that the econometrics employed in the previous papers were deficient.

Also, in their study of the team performance of both exogenously and endogenously formed MBA groups at Harvard, Calder-Wang, Gompers, and Huang (2021), document that the negative effect of diversity is alleviated in cohorts in which teams are endogenously formed. More generally, exogenous board gender-diversification processes are fraught with several limitations (e.g., timing, external influences, factors relating to team dynamics, et cetera). These limitations have contributed significantly to the mixed state of the extant evidence on the impact of board gender diversity. Moreover, behavioral science theory on the relationship between the stages of team development and team performance provides an intuitive explanatory basis for why such limitations might continue to constitute an intricate knot to untie (e.g., see Tuckman, 1965). Worse still, exogenous board gender-diversification processes like explicit quota laws do not practically preclude endogenous selection by the board; they simply mandate it within a limited time frame.

2.2.1.3. Stock price reactions may not represent a value judgment on female directors

In a recent paper, von Meyerinck, et al. (2021) study the 2018 California gender quota law and suggest that the magnitude of the stock price reactions that followed may not represent a firm value-based judgment on female directors. Specifically, although they document large negative announcement returns to the adoption of the gender quota for California firms and large spillover effects for non-California firms, they propose a novel explanation: "Shareholders' disapproval of the government's attempt to legislate non-economic values." They also find that California and non-California firms with higher sensitivity to policy uncertainty react more negatively to the quota's adoption.

The evidence that stock price reactions may not represent an actual value judgment on female directors is also not limited to adverse price reactions. For example, in their study of IPOs, Rau, Sandvik, and Vermalaen (2021) "show that investor preferences for diversity have had a significant effect on the initial returns earned by U.S. firms going public with gender-diverse boards over the past decade." However, finding "no difference in economic fundamentals, such as profitability, between these firms and firms without gender-diverse boards following the IPO," they conclude that "the underpricing [may] reflect recent institutional investor demand for diverse firm boards during the IPO process, possibly [due to] social pressure or an increased focus on corporate social responsibility."

2.2.2. Zero impact on US firm risk-taking and performance closer to reality, on average

There are also reasons why the true impact of gender diversity for US boards might be closer to zero, on average. Again, an inexhaustive list includes (i) tokenism (e.g., see Kanter, 1977) and (related to tokenism) (ii) the snuffing impact of a low individual share of the within-board power/influence distribution (e.g., see Wabara, 2021a).

2.2.2.1. Tokenism

Kanter (1977) popularized that proportions, i.e., the relative numbers of socially and culturally different people in a group, are critical in shaping interaction dynamics. The paper describes four group types based on varying proportional compositions: uniform, skewed, tilted, and balanced. *Uniform* groups have 100% of one type. *Skewed* groups contain a large preponderance (up to 85%) of one type (the numerical dominants) over another (the rare tokens). *Tilted* groups have up 65% of the numerical dominants. *Balanced* groups have no more than 60% of the numerical dominants. Kanter's main argument is that, in skewed groups, the numerical dominants control the group and its culture. It is not until the group reaches a tilted status, with a critical mass of the tokens, that it moves towards less extreme distributions and less exaggerated effects. The latter corresponds to the popularly named critical mass theory (e.g., see Kramer, et al., 2007).

We show, in Figure 2.1, that the average annual proportion of female directors on the boards of publicly listed US firms did not exceed 10% until after the financial crisis in 2008. Since then, although this annual proportion of female directors as a percentage of the board size has risen steadily, it appears to be peaking at around 22%. Accordingly, male directors have been the numerical dominants on US boards, and female directors have been the rare tokens in what has been predominantly skewed boards, gender-wise. The median number of female directors per board per year also did not shift from 1 to 2 until 2016. Hence, to the extent that the critical mass theory is valid, the adverse effects on firm value (documented to coincide with some board gender-diversification events) may well not be the direct result of the mere presence of female directors on the board.

2.2.2.2. Within-group power and influence asymmetry

As Zimmer (1988) points out, Kanter's focus on the importance of the numerical composition follows a strong theoretical tradition within sociology in general (e.g., see Blau, 1977; Homans, 1974; Simmel, 1950) and within race relations research in particular (e.g., see Frisbie and Neidert, 1977; Giles, 1977; Marden and Meyers, 1973). However, the minority relations literature suggests that issues of power, privilege, and prestige are considerably more important than numbers for understanding relations between dominant and subordinate groups (Gittler, 1956; Noel, 1968; Yetman, 1985).

Consistent with both the theoretical tradition in sociology (above) and the minority relations literature, Wabara (2021a), using a unique setting as a pseudo-laboratory to study individual and collective financial risk-taking in action, documents that proportions must lead to sufficient power and influence to be meaningful for the eventual evasion of tokenism or the guarantee of an exit from a token status. Put differently, a low-proportion subgroup, or even just one individual, with an enormous share of the within-group power and influence distribution, can be "the dominant," not "the token."

2.2.3. When board gender diversity can have a positive impact

The implication of the foregoing appears straightforward. Except a study can effectively isolate contexts in which female directors have sufficient power/influence to exert themselves on the board, any direct attribution of negative or positive impact would be puzzling, at least. In effect, Wabara (2021a) documents several related and intuitive empirical results. First, on average, a female will take less excessive financial risks (i.e., mainly when the ex-ante expected profit is zero). However, in small groups, an individual's (or a subgroup's) share of the within-group power and influence distribution, not just numerical strength, determines whether their tendencies manifest in the collective decisions or outcomes of the group.

These and other Wabara (2021a) results are interesting for several reasons. First, they are consistent with the central assumption in Adams, Almeida, and Ferreira (2005).[30] Second, the results suggest a crucial condition (i.e., when the risk-taking is qualitatively excessive) under which female directors might be more risk-averse in a value-maximizing way for the firm. Third, the results also help push the needle in reconciling[31] the mixed extant empirical evidence on the relationship between gender and risk preferences.

[30] Adams, Almeida, and Ferreira (2005) state that executives can only impact firm outcomes if they have influence over crucial decisions. Wabara (2021a) shows that in small groups, the individual share of the within-group power and influence distribution determines whether personal preferences (such as for excessive risk-taking) manifest in the collective decisions and outcomes.

[31] Jianakoplos and Bernasek (1998), using U.S. sample data of household holdings of risky assets, conclude that women are more financially risk-averse; Adams and Funk (2012), using survey data, conclude that female directors are different from women in the general population and may be more risk-loving than male directors. Wabara (2021a) document that every female need not be less risk-loving than any male. However, the difference in risk preferences by gender may relate to whether the risk is qualitatively excessive or not.

Moreover, the results provide a valuable framework for organizing a new study focusing on reconciling the mixed extant evidence on the impact of board gender diversity.

Consequently, we build our hypotheses directly on the Wabara (2021a) results.

Hypotheses:

If a female director is unlikely to have any personal power or influence on the board, her addition to the board will have no significant impact on firm risk-taking and financial performance. However, with increasing power/influence on the board via greater numerical strength or non-token aggregate position, female directors will tend to reduce the excessive risk-taking behavior of the firm and, to the extent that the gender-diversification is non-disruptive, the expected risk-reduction effect can feature a significant increase in profitability and firm value as a component.

The profitability and firm value components are consistent with Berliant and Fujita (2012), who theoretically model diversity as endogenous and immobile but demonstrate how multiple cultures interacting can improve productivity.

2.3. Empirical Model and Examinations

2.3.1. Base samples, variables, and summary statistics

We obtain our base samples of board data for publicly listed firms from the ISS Directors Data and Execucomp. The ISS Directors Data contains information on Board size, the gender of the directors, the number of independent directors, and their positions on the board (including whether they are in non-CEO leadership positions). A non-CEO leader may be, among other things, a lead independent director or a senior independent director. We rely on the Execucomp data mostly to construct and cross-check our CEO-transition variable. We obtain our data on firm fundamentals from Compustat. We intersect the firm fundamentals data from Compustat with the director and board characteristics data from ISS Directors Data and Execucomp. In Appendix 2.1, Table 1, we define all variables used in this study, including the firm and board characteristics.

Table 2.1 summarises the firm and board characteristics from the intersection of the Compustat and ISS Directors Data. We call this Base Sample 1. The mean (median) leverage is 22% (21%), and the mean (median) operating profitability is 14% (13%). The median board has nine directors, of which just one is a female director. The median board has seven independent directors, of which just one is a female independent director. The median board has only one non-CEO leader (e.g., lead director), and the gender of that non-CEO leader is male. On average, only about 3% of the boards have a female CEO or a female chairman. About 1 % of the boards have a chairman/CEO that is female.

Using the same data intersection, we plot, in Figure 2.1, the average proportion and the median number (both per year) of female directors on publicly listed US firms. It shows that following the global financial crisis in 2008, the average ratio of female directors on US boards began to rise steadily from about 10% and now seems to be peaking at about 22%. The median number of female directors per board per year did not shift from one to two until 2016 and has remained unchanged since then. However, the shift occurred before the #Me Too Movement (in 2017). It also happened before the "Big Three" push (in 2017) for more women on the board and before the California gender quota law (in 2018).

Table 2.2 summarises the firm and board characteristics from the intersection of the Compustat and Execucomp data. We call this Base Sample 2. The summary statistics for the firm characteristics are similar to those in Table 2.1. For example, the mean (median) leverage is 21% (20%), and the mean (median) operating profitability is 13% (13%).

2.3.2. Summary of hypotheses development

Following our arguments in the hypotheses development section (i.e., Section 2.2), we concluded with a set of hypotheses. First, we stated that "if a female director is unlikely to have any personal power or influence on the board, her addition to the board will have no significant impact on firm risk-taking and financial performance." In addition, we also stated that "with increasing power/influence on the board (via greater numerical strength or non-token aggregate position), female directors will tend to reduce the excessive risk-taking behavior of the firm and, to the extent that the gender-diversification is non-disruptive, this effect can feature significant increases in profitability and firm value." We also highlighted that our hypotheses are intuitive and consistent with Wabara (2021a), Yetman (1985), Adams, Almeida, and Ferreira (2005), and Berliant and Fujita (2012).

2.3.3. Simplifying our hypotheses

The first part of our hypotheses relates to a board structure in which the female director is unlikely to have any power or influence. To use the terminology of Kanter (1977), she is a rare token in a skewed group. In this case, she would be unlikely to have any effect on firm risk-taking and performance. Formally, we delineate this part as follows:

Hypothesis 1 (H1):

If a female director is unlikely to have any personal power or influence on the board, her addition to the board will have no significant impact on firm risk-taking and financial performance.

The second part of our hypotheses relates to a changing power and influence condition on the board for the female subgroup of directors. It has two subparts. The first subpart continues along the line of Kanter (1977) by increasing the power and influence of the female subgroup of directors on the board through a significant increase in their numerical strength. In this case, the board would advance from a skewed group to a tilted group. To the extent that they vote similarly on risk-related and other matters, the female directors would become more likely to have some aggregate effect on firm risk-taking and performance. Again, formally, we delineate that part of our hypotheses as follows:

Hypothesis 2a (H2a):

With increasing power/influence on the board via greater numerical strength, female directors will tend to reduce the excessive risk-taking behavior of the firm and, to the extent that the gender-diversification is non-disruptive, the expected risk-reduction effect ***can*** *feature a significant increase in profitability and firm value as a component*

The second subpart of the second part of our hypotheses has the same objective of increasing the power and influence of the female subgroup of directors on the board but emphasizes the minority relations literature approach of power, privilege, and prestige (e.g., Gittler, 1956; Noel, 1968; Yetman, 1985). In this case, the female directors will also need to occupy non-token aggregate positions on the board rather than only increase their numerical strength. The result (in terms of increasing the chances that the preferences of the female subgroup directors manifest in the collective board decisions, and ultimately, firm outcomes) would be similar to that of increasing the numerical strength of the female directors; in several ways. Also, formally, we state that part of our hypotheses as follows:

Hypothesis 2b (H2b):

With increasing power/influence on the board via non-token aggregate position, female directors will tend to reduce the excessive risk-taking behavior of the firm and, to the extent that the gender-diversification is non-disruptive, the expected risk-reduction effect ***can*** *feature a significant increase in profitability and firm value as a component.*

As we show later, we further split H2b in two for technical reasons. We call the splits H2b(1) and H2b(2). Overall, the simplified hypotheses are entirely consistent with Wabara (2021a). However, we emphasize the verb "**can**" in hypotheses 2a and 2b because part of our conjecture is that increases in profitability and firm value are unlikely to materialize if other adverse factors (e.g., a disruptive gender-diversification process) intervene.

2.3.4. Hypotheses testing: approach, samples, theoretical model, and process

Although the negative effect of diversity is alleviated in cohorts in which teams are endogenously formed[32], replicating pure or even quasi-experimental conditions in a board setting involving endogenous selections would not be too straightforward. For example, to test H1, one may need previously homogenous male boards to each select an ordinary[33] female director. The just-described event would be the endogenous treated condition per firm. One would then need to know what would happen to each of the boards in the counterfactual situation. For example, what would be the impact on a board's risk-taking decision if the gender of the new ordinary director were male instead? The latter would be the endogenous control condition per firm.

Indeed, in some cases, if no other significant changes occur on the board, then it might seem plausible that the (average) difference between the treated and control conditions (across stacked[34] cohorts of firms) would reflect the actual impact of the female gender. However, such a conclusion might still be rather hasty for the corporate board contexts to which the firms in our base samples belong (i.e., publicly listed US firms). The

[32] E.g., see Calder-Wang, Gompers, and Huang (2021)
[33] We define ordinary to mean unlikely to exert any significant power or influence to alter the board's decision.
[34] As we detail in sections 2.3.4.1. to 2.3.4.3., for each of the simplified hypotheses tests, we define specific events and their quasi-temporal counterfactuals. The events and the counterfactuals might occur for different firms in different years thus creating multiple event year cohorts.

latter is because several ex-ante beliefs and motives could drive the endogenous director selections (e.g., see Hermalin and Weisbach, 1988; Agrawal and Knoeber, 2001; Bertrand and Schoar, 2003), thus complicating the interpretation of the results.

Nevertheless, a quasi-experimental identification strategy holds the promise of dealing with the possibility that the negative female gender effect on firm value documented in Adams and Ferreira (2009) is principally a mere artifact of an inadvertent triggering of intense board oversight. In other words, a quasi-experimental identification strategy would increase the likelihood that some quasi-counterfactual events in which new male directors trigger intense board oversight might be in the set of control samples for events in which new female directors also trigger intense board oversight.

Accordingly, we adopt this quasi-experimental identification strategy and make concerted analytical efforts to rule out several potential ex-ante beliefs and motives for the endogenous director selections. In effect, we incorporate mitigating selection issues into our structuring and automated search (within the Base Samples) of the quasi-treatment and control subsamples for the respective hypotheses. We, below, describe these efforts.

2.3.4.1. Structuring, automated search for treated and control samples, and distribution

2.3.4.1.1. Hypothesis 1 (H1)

The fundamental question that the test of H1 would address is, does the addition of an ordinary female director to a previously homogenous male board affect the risk-taking and financial performance of the firm? We define an ordinary director as a director who is unlikely to unilaterally exert sufficient power or influence to alter the board's collective decision on risk-related and other matters. Our ex-ante expectations are evident from the statement of the hypothesis. We believe not. The next crucial question is, how then do we obtain the quasi-treated and control subsamples for the hypothesis test?

2.3.4.1.1.1. Structuring and automated search for treated and control samples (H1)

The ISS Directors data (used in the construction of Base sample 1) identifies directors in non-CEO leadership roles on the board for any given year. We also know that CEOs—male or female—tend to have large shares of power and influence in firms and corporate boards (e.g., see Adams, Almeida, and Ferreira, 2005; Albuquerque and Miao, 2013). Therefore, introducing a new female director in a non-leading role to a previously homogenous male board with a male CEO would nicely approximate the introduction of an ordinary female director to the board. Similarly, introducing a new male director in a non-leading role to a previously homogenous male board with a male CEO would adequately approximate the introduction of an ordinary male director to the board.

In other words, for a test to evaluate the effect of a female director unlikely to have any significant power/influence on the board, the former would correspond to the treated sample and the latter to the control sample. Both constitute our first set of quasi-temporal counterfactual events. We also impose additional sample search conditions. Overall, and as schematized in Figure 2.2, the sample search conditions are as follows:

1. The boards are homogenously male with male CEOs for at least three years before introducing the new ordinary female director (for the treated firms) or the new ordinary male director (for the control firms). We call this the period from $t-3$ to $t-1$, or in other sections, periods 1, 2, and 3.
2. The first year the new director appears in the sample as a board member is the first of the three years post-event. This characterization is reasonable because new directors typically become part of the board before the end of the fiscal year that they first appear in the sample as board members. We call this period t, or in other sections, period 4.
3. The board size only increases by one (i.e., the new director).

4. The new directors stay for at least another two years, during which the board size/composition remains unchanged. We call this the period from $t + 1$ to $t + 2$, or in other sections, periods 5 and 6.
5. The male CEOs are unchanged the three years before and the three years after.
6. A firm stays in the treated (control) sample if there is at least one firm in the same Fama-French 48 industry in the control (treated) sample. To verify that our results remain qualitatively unchanged, we drop this condition in the robustness test.

Ultimately, to enhance the chances of a credible impact (if any) attributability to the female director, we use three-year pre-/post-event windows. The three-year half-windows ensure that the time horizon is neither too short (to not include the direct impact, if any, of the new director) nor too long (for the effect of the new director to be confounded by other events). Accordingly, we execute the automated search on Base sample 1.

2.3.4.1.1.1.1. Annual distribution of the treated and control samples (H1)

We show, in Figure 2.3, the yearly distribution of the treated and control samples for hypothesis test 1. Column 1 shows the fiscal year ends represented in Base Sample 1. Column 2 shows the number of firm-year observations per fiscal year in the base sample. Columns 3 and 4 show the results of the automated treated and control samples search. Specifically, Column 3 shows the proportion of the firm years that are in the control sample. On average, about 3.8% of the 8,663 observations are in the control sample. The majority of the control sample falls in the period between 2008 and 2016.

Similarly, column 4 shows the proportion of the firm years that are in the treated sample. On average, about 7% of the 8,663 observations are in the treated sample. The majority of the treated sample also falls in the period between 2008 and 2016. Overall, the results suggest that following the 2008 financial crisis, previously homogenous male boards with male CEOs added at least one new director to the board in non-leading roles. However, of those new directors added, nearly twice as many were female directors. This interpretation is also consistent with the trend shown earlier in Figure 2.1.

Notably, most of these endogenous board re-composition events happened before the #Me Too Movement (in 2017), before the "Big Three" push (in 2017) for more women on the board, and before the California gender quota law (in 2018).

2.3.4.1.2. Hypothesis 2a (H2a)

The fundamental question that the test of H2a would address is, does the increase in the power and influence of female directors on the board (via greater numerical strength for the female subgroup) affect the risk-taking and financial performance of the firm? Our ex-ante expectations are also evident from the statement of the hypothesis. We believe yes, particularly concerning excessive risk-taking (consistent with Wabara, 2021a). We believe so as long as the gender-diversification process is non-disruptive to the firm's operations. The vital question that then follows is, how do we obtain the quasi-treated and control subsamples for the hypothesis test? We detail this below.

2.3.4.1.2.1. Structuring and automated search for treated and control samples (H2a)

To generate the treated and control samples for this test, we build on the automated sample search structure for hypothesis 1. First, we restrict Base Sample 1 to gender-skewed boards with a male CEO and only one female director for at least the past three years. This restriction is equivalent to keeping only boards previously treated at some point in the manner described in hypothesis 1. We impose this restriction for a couple of reasons.

The first is that a board already having a female director and then adding more would ensure that we capture the effect of an increase in numerical strength and not that of the first time addition of a female director. The second is that it would help mitigate the impacts of ex-ante endogenous director selection beliefs or motives. For example, suppose the endogeneity issue is that, on average, female directors self-select into (or are chosen by) previously homogenous male boards that tend to be more aggressive with the reduction of excessive risk. In that case, we would still be looking at largely similar boards pre-(H2a)-event. The event being the addition of more ordinary female directors. Consequently, and as schematized in Figure 2.4, our sample search conditions for this test are as follows:

1. The boards are skewed majority male with male CEOs and only one female director for at least three years before introducing an additional one or more ordinary female directors (for the treated firms) or ordinary male directors (for the control firms). Again, we call this the period from $t-3$ to $t-1$, or in other sections, periods 1, 2, and 3.
2. The first year in which the new (additional one or more) ordinary directors appear in the sample as board members is the first of the three years post-event. Again, this characterization is reasonable because new directors typically become part of the board before the end of the fiscal year that they first appear in the sample as board members. We also call this period t, or in other sections, period 4.
3. The board size only increases by the number of new directors added.
4. The new directors stay for at least another two years, during which the board size/composition remains unchanged. We also call this the period from $t+1$ to $t+2$, or in other sections, periods 5 and 6.
5. The male CEOs are unchanged the three years before and the three years after.

6. A firm stays in the treated (control) sample if there is at least one firm in the same Fama-French 48 industry in the control (treated) sample. To verify that our results remain qualitatively unchanged, we drop this condition in the robustness test.

To enhance the chances of a credible impact (if any) attributability to the female directors, we use three-year pre-/post-event windows. The three-year windows ensure that the time horizon is neither too short (to not include the direct impact, if any, of the new directors) nor too long (for the effect of the new directors to be confounded by other events). Accordingly, we execute the automated search on Base sample 1.

2.3.4.1.2.1.1. Annual distribution of the treated and control samples (H2a)

We show, in Figure 2.5, the yearly distribution of the treated and control samples for hypothesis test 2a. Column 1 shows the fiscal year ends represented in Base Sample 1. Column 2 shows the number of firm-year observations per fiscal year in the base sample. Columns 3 and 4 show the results of the automated treated and control samples search. Specifically, Column 3 shows the proportion of the firm years that are in the control sample. On average, about 4.5% of the 8,663 observations are in the control sample. The majority of the control sample falls in the period between 2007 and 2016.

Similarly, column 4 shows the proportion of the firm years that are in the treated sample. On average, about 5.4% of the 8,663 observations are in the treated sample. The majority of the treated sample falls in the period between 2009 and 2016. Overall, the results suggest that following the 2008 financial crisis, the previously skewed majority male boards with male CEOs and only one female director began to add at least one new director to the board in non-leading roles. However, of those new ordinary directors added by such boards, a majority of them were female directors. This interpretation is also consistent with the trend shown earlier in Figure 2.1.

Notably also, the bulk of these endogenous board re-composition events happened before the #Me Too Movement (in 2017), before the "Big Three" push (in 2017) for more women on the board, and before the California gender quota law (in 2018).

2.3.4.1.3. Hypothesis 2b (H2b)

The fundamental question that the test of hypothesis 2b would address is whether the increase in the power and influence of female directors on the board (via more significant non-token aggregate position for the female subgroup) affects the risk-taking and financial performance of the firm? Again, our ex-ante expectations derive from the statement of the hypothesis. We believe yes, particularly concerning excessive risk-taking (consistent with Wabara, 2021a). Again, we believe so as long as the gender-diversification process is non-disruptive to the firm's operations. The question that then follows is, how do we obtain the quasi-treated and control subsamples for the hypothesis test?

We note that the position of CEO is one of the most influential on the board, especially so for publicly listed US boards (e.g., see Adams, Almeida, and Ferreira, 2005; Albuquerque and Miao, 2013). Consequently, to test this hypothesis, we consider situations in which a firm changes CEO, replacing a male CEO with a female CEO. In this case, the control sample would consist of quasi-counterfactual samples in which a firm changes CEO, replacing a male CEO with another male CEO. However, to capture the actual effect of the greater positional power for the female subgroup of directors when the CEO changes from male to female, it may not suffice to simply compare boards with male-to-female CEO transitions to those with male-to-male CEO transitions. Additional care would be more appropriate in structuring the treated and control sample.

For example, we would need to exclude all boards with one or more female directors from the control sample since, from hypothesis 2a, we expect such boards (even with a male CEO) to have a reduced tendency for excessive risk-taking. To demonstrate the latter, we split hypothesis 2b in two. Let H2b(1) be the sub-hypotheses in which we exclude, from the control sample, firms with male-to-male CEO transitions but whose boards include one or more female directors. Let also H2b(2) be the sub-hypotheses in which we have, in the control sample, firms with male-to-male CEO transitions and whose boards have one or more female directors. In those cases, the following details follow:

2.3.4.1.3.1. Structuring, automated search for treated and control samples (H2b(1))

As schematized in Figure 2.6, our sample search conditions for this test are as follows:

1. The treated board has a unique male (i.e., not a joint) CEO and may have one or more female directors for at least three years before transitioning from male CEO to female CEO. The control board also has a unique male (i.e., not a joint) CEO but no female director for at least three years before transitioning from the male CEO to a new male CEO. We call this the period from $t-3$ to $t-1$, or in other sections, periods 1, 2, and 3.
2. The first year in which the CEO transitions occur in the sample is the first of the three years post-event. Again, this characterization is reasonable because new directors typically become part of the board before the end of the fiscal year that they first appear in the sample as board members. We also call this period t, or in other sections, period 4.
3. The board size does not increase by more than one (to accommodate the event that the new CEO was an outsider).

4. The new CEO stays for at least another two years, during which the new board size/composition remains unchanged. We also call this the period from $t + 1$ to $t + 2$, or in other sections, periods 5 and 6.
5. A firm stays in the treated (control) sample if there is at least one firm in the same Fama-French 48 industry in the control (treated) sample. To verify that our results remain qualitatively unchanged, we drop this condition in the robustness test.

To enhance the chances of a credible impact (if any) attributability to the female directors, we use three-year pre-/post-event windows. The three-year windows ensure that the time horizon is neither too short (to not include the direct impact, if any, of the new directors) nor too long (for the effect of the new directors to be confounded by other events). Accordingly, we execute the automated search on Base sample 2.

2.3.4.1.3.1.1. Annual distribution of the treated and control samples (H2b(1))

We show, in Figure 2.7, the yearly distribution of the treated and control samples for hypothesis test 2b(1). Column 1 shows the fiscal year ends represented in Base Sample 2. We restrict the firm years for Base Sample 2 to 2003-2019 (to remain broadly comparable to Base Sample 1, in which the fiscal firm years range from 2006 to 2019). Column 2 shows the number of firm-year observations per fiscal year in the base sample. Columns 3 and 4 show the results of the automated treated and control samples search. Specifically, Column 3 shows the proportion of the firm years that are in the control sample. On average, about 11.6% of the 15,231 observations are in the control sample. The majority of the control sample falls between 2003 and 2016, with a peak in 2006.

Similarly, column 4 shows the proportion of the firm years that are in the treated sample. On average, about 3.1% of the 15,231 observations are in the treated sample. The majority of the treated sample falls between 2005 and 2015, with a peak in 2009. Overall, the results suggest that following the 2008 financial crisis, the bulk of CEO transitions for publicly listed US firms have been male-to-male, but a switch in the gender of CEOs from male to female appears to have picked up considerably in 2009. Nevertheless, these US firms with homogenous male boards remain 4x as likely to replace a male CEO with a male CEO as a gender-skewed majority male board replacing a male CEO with a female CEO.

Notably, the bulk of these CEO transition events happened before the #Me Too Movement (in 2017), before the "Big Three" push (in 2017) for more women on the board, and before the California gender quota law (in 2018).

2.3.4.1.3.2. Structuring and automated search for treated and control samples (H2b(2))

As schematized in Figure 2.8, our sample search conditions for this test are as follows:

1. The treated board has a unique male (i.e., not a joint) CEO and may have one or more female directors for at least three years before transitioning from male CEO to female CEO. The control board also has a unique male (i.e., not a joint) CEO and may also have one or more female directors for at least three years before transitioning from the male CEO to a new male CEO. We call this the period from $t-3$ to $t-1$, or in other sections, periods 1, 2, and 3.
2. The first year in which the CEO transitions occur in the sample is the first of the three years post-event. Again, this characterization is reasonable because new directors typically become part of the board before the end of the fiscal year that they first appear in the sample as board members. We also call this period t, or in other sections, period 4.
3. The board size does not increase by more than one (to accommodate the event that the new CEO was an outsider).

4. The new CEO stays for at least another two years, during which the new board size/composition remains unchanged. We also call this the period from $t + 1$ to $t + 2$, or in other sections, periods 5 and 6.
5. A firm stays in the treated (control) sample if there is at least one firm in the same Fama-French 48 industry in the control (treated) sample. To verify that our results remain qualitatively unchanged, we drop this condition in the robustness test.

To enhance the chances of a credible impact (if any) attributability to the female directors, we use three-year pre-/post-event windows. The three-year windows ensure that the time horizon is neither too short (to not include the direct impact, if any, of the new directors) nor too long (for the effect of the new directors to be confounded by other events). Accordingly, we execute the automated search on Base sample 2.

2.3.4.1.3.2.1. Annual distribution of the treated and control samples (H2b(2))

We show, in Figure 2.9, the yearly distribution of the treated and control samples for hypothesis test 2b(2). Column 1 shows the fiscal year ends represented in Base Sample 2. Column 2 shows the number of firm-year observations per fiscal year in the base sample. Columns 3 and 4 show the results of the automated treated and control samples search. Specifically, Column 3 shows the proportion of the firm years that are in the control sample. On average, about 25.6% of the 15,231 observations are in the control sample. The majority of the control sample falls between 2003 and 2016, with a peak in 2006.

Similarly, column 4 shows the proportion of the firm years that are in the treated sample. On average, about 3.1% of the 15,231 observations are in the treated sample. The majority of the treated sample falls between 2005 and 2015, with a peak in 2009. Overall, the results suggest that following the 2008 financial crisis, the bulk of CEO transitions for publicly listed US firms have been male-to-male, but a switch in the gender of CEOs from male to female appears to have picked up considerably in 2009. Nevertheless, these US firms with gender-skewed majority male boards remain over eight times (8x) as likely to replace a male CEO with a male CEO as to replace a male CEO with a female CEO.

Notably also, the bulk of these CEO transition events happened before the #Me Too Movement (in 2017), before the "Big Three" push (in 2017) for more women on the board, and before the California gender quota law (in 2018).

2.3.4.2. Measuring firm-level risk-taking, firm value, and gauging market's reaction

Board directors can influence managerial choices and firm policies through their votes on strategic corporate matters (e.g., see Forbes and Milken, 1999; Monks and Minnow, 1996). They may also do so through their formal work on the monitoring and advising committees, where they may use different forms of incentives and strategery (e.g., see Song and Thakor, 2006; Adams and Ferreira, 2007; Edmans, Gabaix, and Landier, 2009; Bebchuk and Fried, 2010; Faleye, Hoitash, and Hoitash, 2011). Also, firm policies more directly reflect senior leaders' preferences than volatility because other factors outside the managers' control can influence volatility (e.g., see Pan, Siegel, and Wang, 2017). Consistent with this view, we first use firm-level policy-related risk determinants (or indicators) to measure female directors' impact on the corporate risk-taking culture.

We use five firm-level risk-taking proxies widely used in the literature. The first is book leverage, defined as the ratio of total book debt to total book assets (e.g., see John, Litov, and Yeung, 2008). While leverage is not necessarily a bad thing (e.g., see Palepu, 1990; Kaplan, 1994), excessive leverage can heighten the probability of firm bankruptcy and lead to the imposition of significant constraints on a firm's ability to make long-term value-enhancing investments (e.g., Warner, 1977; Titman, 1984). Indeed, our hypotheses hinge on the idea that only with sufficient power and influence on the board (either via greater numerical strength or via non-token aggregate position) can female directors influence boards' collective decisions. However, consistent with Wabara (2021a) and Berliant and Fujita (2012), the direction of that impact will be to reduce excessive risk (if that does exist) and enhance firm value. Unfortunately, reducing leverage alone will not suffice. In some cases, it could even be that an increase in leverage might be required.

Interestingly, cash holdings offer financial flexibility and lower risk (e.g., see Pan, Siegel, and Wang, 2017). Hence, we include cash/asset (i.e., the ratio of book cash holdings to total book assets) to our list of proxies. However, a transient increase in the cash/asset ratio may not always reflect better financial stability for the firm. To measure to what extent an increase in the cash/asset ratio reflects financial soundness, we include a related proxy, the volatility of cash/asset, which is the standard deviation of the ratio. We compute this standard deviation over the previous eight quarters (or two years), consistent with and within our treated and control samples search half-windows of three years. Considered together, an increase in cash/asset accompanied by a decrease in its volatility might also indicate relative reductions in the probability of bankruptcy.

We also include two other firm-level risk-taking proxies, namely, tangibility and operating profitability. Tangibility is book property, plant, and equipment (PPE) scaled by the total book assets, and operating profitability is operating income before depreciation scaled by total book assets. We include both proxies because the non-reduction/increases in tangibility might indicate relative reductions in the expected costs of financial distress in the event of default. Similarly, relative increases (decreases) in operating profitability might be indicative of relative reductions (increases) in the underlying business risks, respectively (e.g., see Aggarwal et al., 2010).

As Pan, Siegel, and Wang (2017) document, volatility may not always reflect the actual impact of female directors on firm-level policy. However, considered together with the firm-level proxies, it might help uncover the overall direction of the effects. Hence, we use market-level volatility to gauge the market's reaction to (or its anticipation of) the level of the risk-taking inherent in the firm. Specifically, we use the volatility (i.e., standard deviation) of the market-adjusted return. The market-adjusted return is the stock return less same-period return on the CRSP value-weighted portfolio of NYSE/AMEX/Nasdaq stocks. We compute this standard deviation over 36 preceding months, again consistent with and within our treated and control samples search half-windows of three years.

Unfortunately, in addition to the real possibility of disconnects between firm-level events such as the addition of female directors and stock market reactions, the latter may not happen efficiently (e.g., see Fama, 1970). For example, intense news and public communication (i.e., buzz) about a non-fundamental firm-level event[35] can send the market reaction swinging, potentially increasing return volatility. In contrast, a hush about an otherwise important firm-level event may not generate a commensurate reaction from the market. In our context, given the period that the majority of our treated and control samples cover (i.e., 2003 to 2015), we do not expect the event of adding ordinary female directors (to male-dominated boards with male CEOs) to generate sufficient buzz that will significantly impact the market-adjusted return volatility. Today, however, given the various push and laws to have more female directors on US boards, similar events could generate a different market reaction. Whether the market reaction would indeed be about the presence of female directors is a different matter. In either period, nonetheless, we would expect CEO transition events to generate sufficient news to impact the market.

[35] One might understandably put the 2020 market actions surrounding the GameStop stock in this category.

Finally, to measure firm value or performance, we stylistically follow the popular trend in the corporate governance literature by using Tobin's q, defined as total book assets minus the book value of equity plus the market value of equity, divided by the book value of total assets (e.g., see Adams and Ferrera, 2009). We do this for two reasons. The first is for direct comparability with that literature. The second is that this definition of Tobin's q precisely defines market-to-book in Rajan and Zingales (1995), the implications of which we discuss in detail in a few subsections below.

Overall, given that individual changes in these proxies can mean several things, we focus not on any particular proxy alone but the general consistency of the complete story their changes collectively tell in response to each of the events in our hypotheses tests.

2.3.4.3. Implementing the hypotheses tests using stacked difference-in-differences (DiD)

Typically, a standard DiD implementation uses a single event date around which the generic DiD computations flow.[36] However, as the preceding sections show, one of our primary innovations in the study of board gender diversity is that we define specific treatment events and their quasi-temporal counterfactual events for each of our simplified hypotheses tests. These treatment events and their counterfactuals occur in different years for different firms, creating multiple event-year cohorts per hypothesis test.

For example, a firm, say T^{1}_{2009}, might be treated in the manner described for H1 in 2009. Given our three-year pre-event and three-year post-event symmetric half-windows, the treated sample for the firm will run from 2006 to 2011. The 2006 to 2008 observations will correspond to the pre-event period, while the 2009 to 2011 observations will correspond to the post-event period. Accordingly, firms in the control or counterfactual sample would satisfy the stated search conditions with respect to the treatment event year.

[36] For an overview of DiD methodology, challenges and adaptations, see Baker, Larker, and Wang (2021).

In that case, all firms treated in the same year (i.e., all T_{2009}^{i}) and all firms that qualify as the relevant controls (i.e., all C_{2009}^{i}) will belong to the same (2009) cohort. A typical DiD implementation would then be based on taking the conventional averages and differences around the single cohort event date of 2009. However, in this study's context, another set of firms, T_{t}^{i}, might be treated in year t for the same test for H1. Those firms will also have their set of controls, C_{t}^{i}, all of which will belong to a different cohort t. All these will lead to several event dates or cohorts of counterfactual samples. In effect, there are nine (9) and twelve (12) cohorts in our Base Sample 1 and Base Sample 2, respectively.

One solution in the literature for dealing with multi-cohort events is to use the staggered DiD implementation. However, this approach does not fit well with our sample structuring using fixed event windows. Moreover, Baker, Larker, and Wang (2021) show that staggered DiD designs are likely to be biased in the presence of treatment effect heterogeneity. Indeed, our treatment and control samples vary in multiple dimensions. For example, Figure 2.10 shows the heterogeneities in the distributions (by Fama-French industry classification) of the baseline treated and control samples for each of our simplified hypotheses tests. Specifically, each of the four hypothesis tests involves close to twenty (20) different industries. Thus, an alternative and more robust approach would be preferable. Baker, Larker, and Wang (2021) also document that one such robust alternative approach is the stacked DiD regression estimator, as has been used by Cengiz et al. (2019) on event-specific datasets defined over a fixed time window.[37]

---

[37] Historically, stacking as an empirical identification process is not new. It has long been in use in seismic engineering and geophysical imaging as a robust empirical strategy for identifying complex subsurface objects. One good example is the common midpoint (CMP) stacking process introduced by William Mayne in 1962. CMP stacking means "the summation of a collection of seismic traces from different records into a single trace. It can be considered the simplest way to improve the signal-to-noise ratio (SNR) in prestack seismic data processing. It can help to quickly obtain a meaningful post-stack seismic image without wavefield continuation." For more details on CMP stacking and similarities to its implementation in Finance, see https://reproducibility.org/RSF/book/uh/avostack/paper_html/node3.html.

The stacked DiD approach over fixed event window fits our structured quasi-temporal event counterfactual samples better. For example, our events are all defined over a fixed time window, $W = \{t-3,\ t-2,\ t-1,\ t,\ t+1,\ t+2\}$.[38] Like in Cengiz et al. (2019), our stacked DiD approach corresponds to a setting where the events happen contemporaneously. We also avoid using past treated units as effective comparison units, which may occur with a staggered design. Nevertheless, unlike in Cengiz et al. (2019), our fixed time window is symmetric and shorter. Hence, we call our approach a multi-cohort stacked DiD over a fixed, short, symmetric event window. The exact event dates for our simplified hypotheses are in the interval between $t-1$ and t, presumably the midpoint.[39]

2.3.4.3.1. Overview of multi-cohort stacked DiD over a fixed, short, symmetric event window

Figure 2.IA.1 (in Appendix 2.IA) shows that, in the absence of time fixed effects, the multi-cohort stacked DiD analysis is equivalent to a single-cohort (or single date) stacked DID analysis involving multiple units. To quickly illustrate this, let a vertical stack mean averaging across units, as follows:

$$RI_{\omega,\ Treated}^{Stacked} = 1\{W = \omega,\ Treated = 1\} * \frac{1}{N_1}\sum_{i=1}^{N_1} RI_i \quad (1)$$

$$RI_{\omega,\ Control}^{Stacked} = 1\{W = \omega,\ Treated = 0\} * \frac{1}{N_0}\sum_{i=1}^{N_0} RI_i \quad (2)$$

Other empirical papers that have used various adaptations of the fixed event time window stacked DiD include Gormley and Matsa (2016), Gormley, Matsa, and Milbourn (2013), and Gormley and Matsa (2011).

[38] The half-window, $W^- = \{t-3,\ t-2,\ t-1\}$ contains the three discrete dates in the pre-event period. The half-window, $W^+ = \{t,\ t+1,\ t+2\}$ contains the three discrete dates in the post-event period. $t-1$ is the date just before the event, and t is the first date that any differential effects associated with the counterfactual events may materialize.

[39] Our approach draws significant inspiration from the CMP methodology used in seismic engineering and geophysical imaging. The same is true for some of the languages we employ in parts of our analyses.

Where RI is any risk indicator or proxy variable of interest. N_1 and N_0 are the number of units (firms, in the context of this study) in the treated and control groups, respectively. $W = \{t-3,\ t-2,\ t-1,\ t,\ t+1,\ t+2\}$. $Treated = \{0, 1\}$ where one (1) corresponds to a treated sample and zero (0) corresponds to a control sample.

In both types of DiD analyses, the treated and control samples from different firms are pooled according to their treatment conditions (i.e., treated group or control group) and aligned by their event date. Specifically, in the single-date or single-cohort DiD analysis involving multiple firms, the event date is the same for all firms. The observations positioned at the same relative position from the event date have the same year of observation (y.o.b). Hence, the vertical stack of the treatment or control group will take the average of data points having the same y.o.b. The latter will yield vertical stacks of treated and control groups that exhibit parallel trends without time distortions.

However, in the multi-cohort stacked DiD analysis, the pooled and aligned firms belong to different cohorts with different event dates. In this case, the observations at the same relative position from the event date do not have the same y.o.b. Hence, this will yield vertical stacks of treated and control groups that can exhibit parallel trends with severe time distortions if time fixed effects are present.

Also, to quickly review when and how our multi-cohort stacked DiD over a fixed, short, symmetric event window robustly recovers the true image of the DIDs for each of our quasi-temporal counterfactual samples, we consider the following analytical model.[40]

[40] The analytical model also serves as the basis for several simulations we run to gain pictorial insight into our empirical strategy. We collect all modeling and simulation results in the Internet Appendix.

2.3.4.3.1.1. A simple analytical model in which all heterogeneities are due to fixed effects

Suppose there exist true data for an arbitrary set of quasi-temporal counterfactual events[41] defined as follows. For all discrete observation dates, ω, in the event window, W, the true level of a risk indicator is $y^{w,T}$, where $T = \{0, 1\}$ is the set of treatment conditions such that $T = 1$ corresponds to the treated group and $T = 0$ to the counterfactual or control group. Suppose further that whenever a sample of observations drawn from a firm satisfies either treatment condition (i.e., treated/control), the real differences between the measured levels of such observations and those of the true data are due mainly to the full suite of fixed effects associated with the firm, its cohort and industry, and the time (e.g., year) of the observations. Then, symbolically:

$$y_{ijct}^{w,T} = y^{w,T} + m_i + m_j + m_c + m_t + m_{ij} + m_{ic} + m_{it} + m_{jc} + m_{jt} + m_{ct} + m_{ijc} + m_{ijt} + m_{ict} + m_{jct} + m_{jct} + m_{ijct} + \varepsilon_{ijct}^{w,T} \qquad (3)$$

Where $y_{ijct}^{w,T}$ is the observed level of the risk indicator measured at firm i in industry j and cohort c at time t (corresponding to the relative observation date w) with treatment condition T. $y^{w,T}$ is the true level of the risk indicator for a quasi-counterfactual sample with treatment condition T at relative observation date w. m_p is the p fixed-effect and p is the full suite of individual or interactive fixed effects.

[41] Recall that our quasi-temporal counterfactual events represent a pair of defined counterfactual conditions in the post-event period. Take the first of our four hypotheses in the main text (i.e., H1) as a case in point. All boards are homogenous-male boards, each with a male CEO in the pre-event period, $W^- = \{t-3,\ t-2,\ t-1\}$. However, in the post-event period, $W^+ = \{t,\ t+1,\ t+2\}$, the defined counterfactual conditions are as follows. For one group of the boards, each board adds a female director in a non-leading role, all else equal, and for the other group of boards, each board adds a male director in a non-leading role, with all else unchanged. The first group is equivalent to the treatment (or treated) group, and the second is the counterfactual (or control) group. Ideally (e.g., as in H1, H2a, and H2b(2)) but not compulsorily (e.g., as in H2b(1)), the quasi-temporal events would be indistinguishable in the pre-event period.

Proposition 1 (A multi-cohort stacked DiD approach on a set of quasi-temporal counterfactual samples drawn from a fixed, short, symmetric event window can robustly recover the true image of the DiD): *If there are numerous firms per cohort of quasi-temporal counterfactual events, numerous cohorts per industry, and numerous industries in the data sample, all of which are drawn from a fixed, short, symmetric event window, then the simple DiD,* $\overline{RI}^{DiD}$*, computed on the vertical stacks robustly estimates the true DiD. Moreover, suppose the vertical stacks of the anti-symmetry of fixed effects around the event dates and the residual (measurement) error are precisely zero. In that case, the estimated DiD on the vertical stacks of the treated and control samples is exact.*

Proof: *See the designated Internet Appendix (section 2.IA.2.)*

The proof of the proposition follows from systematic manipulations of equation (6). We note that under the sample conditions stated in Proposition 1, $\overline{RI}^{DiD}$, is equivalent to taking a simple DiD on the stacks computed using equations (1) and (2) above.[42]

---

[42] In the Internet Appendix, we also use simulations to provide additional pictorial insights into the proof of Proposition 1. For example, we illustrate pictorially that, in the presence of time fixed effects, the vertical stack of the proxy variable for the treatment/control groups will yield images that exhibit parallel trends with dynamic (or time) distortions. The severity of the time distortions depends on the distribution and magnitude of the amplitudes of the time fixed effects. We further illustrate how firm, cohort, industry, and other fixed effects affect the shape of the vertical stacks in a multi-cohort stacked DiD analysis over a fixed, short, symmetric event window. Specifically, the industry and cohort fixed effects, by themselves, do not distort the image of the underlying counterfactual data. They only bulk-shift the data. Unlike the impact of the time fixed effects, that of the firm fixed effects is to induce a cosmetic separation between the images of the stacked treated and control data only in the vertical sense. The impact of the firm fixed effects can be easily eliminated by normalizing the vertical stacks. To compute the normalized vertical stack of the treated/control samples, first, we calculate the half-window averages of the vertical stacks ($\overline{RI^{Stacked}_{\omega\epsilon W^{-},\ Treated}}$; $\overline{RI^{Stacked}_{\omega\epsilon W^{+},\ Treated}}$) for the treated group and ($\overline{RI^{stacked}_{\omega\epsilon W^{-},\ Control}}$; $\overline{RI^{Stacked}_{\omega\epsilon W^{+},\ Control}}$) for the control group. Finally, we adjust the stacks by the difference between $RI^{Stacked}_{\omega\epsilon W^{-},\ Treated}$ and $RI^{stacked}_{\omega\epsilon W^{-},\ Control}$. The goal is to reduce the average pre-event window difference to zero. The advantage is that the average post-event window difference between the treated and control groups easily corresponds to the estimated DiD, $\overline{RI}^{DiD}$, on the vertical stacks. We further show that when the distortions from the time fixed effects are not severe, the image of normalized vertical stacks nicely mimics the image of the true data over the event window, W.

2.3.4.3.2. Functional regression model for computing the multi-cohort stacked DiD

A challenge in studying the impact of board gender diversity (on publicly listed US firms' financial performance and value) using quasi-temporal counterfactual samples is the non-numerosity of female directors on the board. The latter puts an upper limit to the number of firms, cohorts, or even industries in the treated samples for specific structured events. We mitigate much of the data challenges by keeping our event window short (i.e., three years pre-event and three years post-event) and making "credible impact attributability" the central theme of our structured sample search within the Base Samples.

Next, we adapt a functional regression model for a robust estimation of the multi-cohort stacked DiDs on our publicly listed US board recomposition events. Building on the intuition from our analytical modeling, we adapt the Bertrand and Mullainathan (2003) functional model to directly compute stacked DiDs on each risk indicator for this study. Specifically, we add firm and interactive industry-year fixed effects[43] to a streamlined version of the Bertrand and Mullainathan model, as follows:

$$RI_{i,j,c,t} = \beta * Treated\_Post_{i,j,c,t} + p_t + m_c + \alpha_i + \vartheta_{j,t} + u_{i,j,c,t} \qquad (4)$$

Where β captures the stacked difference-in-differences effect on each of the risk indicators ($RI_{i,j,c,t}$, described above, computed over the 3-year pre-event and 3-year post-event estimation windows for a firm i, in industry j, treated in year t, and belongs to cohort c). $Treated\_Post$ is the interaction of the treated and post dummy variables and p_t, m_c, α_i,

[43] Our baseline industry distribution initially excludes utilities, banks, insurance, and other financial industries. However, in the robustness tests, we add back those industries to check whether our data will be affected qualitatively. We do the latter because we expect the interactive industry-year fixed effects to account for much of the heterogeneities across those dimensions

and $\vartheta_{j,t}$ are the year, cohort, firm, and industry-year fixed effects, respectively, that help minimize the heterogeneities across those dimensions.[44]

2.3.4.4. Other considerations

Given the desirability of greater sample length in studies of this nature, it is helpful that we use the stacked regression estimator, believed to be more robust than the staggered DiD approach (again, see Cengiz et al., 2019; Baker, Larcker, and Wang, 2021). Another potentially valuable part of our empirical methodology is using (as proxies for firm-level risk-taking) variables with well-studied stylized relationships in the capital structure literature. For example, Rajan and Zingales (1995) show that especially for US firms:

$$Lev_i = \alpha + \beta_1 * Tang_i + \beta_2 * TobinsQ_i + \beta_3 * Profitability_i + \varepsilon_i \qquad (5)$$

$$\begin{cases} \beta_1 > 0 \\ \beta_2 < 0 \\ \beta_3 < 0 \end{cases}$$

Where, for a firm i, Lev_i, $Tang_i$, $TobinsQ_i$, $Profitability_i$ are the firm's leverage, tangibility, Tobin's Q, and profitability, respectively, with similar definitions as in this study.

One potential upside is as follows. Suppose that our stacked DiD implementation estimates that the impact of gender diversity, say for hypothesis 2a, has significantly decreased leverage. In that case, a corresponding stacked DID estimate of a decrease in the tangibility and increases in Tobin's q and profitability would be consistent with Rajan

[44] In the Internet Appendix, we also show that this model adaptation produces highly robust empirical estimates. That is: $\beta_{SD} = \overline{RI}_{SD}^{DID}$, where SD refers the simulation data.

and Zingales (1995) and might increase our confidence in our estimation procedure. However, a potential risk/reward part would relate to the interpretation of our results.

For example, Baker and Wurgler (2002) suggest an “equity market timing” theory of capital structure in which the relationship between leverage and Tobin’s Q flows through net equity issues. In effect, they write an equation dividing the change in leverage (ΔLev_t) into net equity issues (NEI_t), change in retained earnings (ΔRE_t), and a function of asset growth ($f(AssetGrowth_t)$). We summarize that equation as follows:

$$\Delta Lev_t = -NEI_t - \Delta RE_t + f(AssetGrowth_t) \qquad (6)$$

Where net equity issue is the change in book equity minus change in balance sheet retained earnings, divided by total assets. Change in (or newly) retained earnings is the change in balance sheet retained earnings, divided by total assets, and asset growth is the change in total assets, divided by lagged total assets.

Their main point is that higher Tobin’s q is associated with higher net equity issues but not strongly related to retained earnings, which (as they argue) rules out the possibility that Tobin’s q affects leverage because it forecasts earnings. They also note two other interesting patterns from their results: (i) the effect of profitability on changes in leverage arises primarily because of retained earnings. Profitable firms issue less equity, but this effect is more than offset by higher retained earnings so that the net impact of higher profits is to reduce leverage. (ii) Firm size plays an essential role at the time of the IPO.

To outline some potential implications, we continue with our previous example. Suppose that our stacked DiD implementation estimates that the impact of gender diversity, say for hypothesis 2a, significantly affects leverage, tangibility, Tobin’s q, and profitability, all in a manner consistent with Rajan and Zingales (1995). A downside would then be to find that the same effects are coincident with increases in net equity issues, in which case the question would be whether firms simply add women to the board during equity market fundraising events. The latter, although curious, may not be farfetched, especially since Rau, Sandvik, and Vermaelen (2021) find that gender diversity boosts under-pricing in IPOs due to higher institutional investor demand for diverse boards.

However, another upside would be that the same gender diversity effects are not coincident with increases in net equity issues, in which case we would examine other channels relating to firm-level managerial actions. For example, a central goal of financial executives in different countries appears to be guaranteeing the financial stability of their company and the availability of funds (e.g., see Stonehill, 1975). Firms, especially when facing liquidity crises, might sell assets to pay down debt (e.g., see Strebulaev, 2007). The dynamic pecking order theory predicts that firms are more concerned about excessively high leverage than excessively low leverage. Profitable firms and firms with greater cash balances are less likely to use external financing (e.g., see Leary and Roberts, 2005).

2.4. Results

2.4.1. Does tokenistic board gender diversity affect firm risk-taking and performance?

Precisely, what happens to the five firm-level risk-taking proxies, the firm value proxy, and the volatility of the market-adjusted return when previously homogenous male boards with a male CEO add a new female director to the board in a non-leading role? Our ex-ante expectations derive from the first part of our hypotheses (i.e., hypothesis 1 or H1). Hence, we do not expect any statistically significant changes in the firm-level risk proxies. We also do not expect any statistically significant changes in the firm value proxy or the volatility of the market-adjusted return. More so for the latter because adding just one ordinary female director to the board would not have created significant market communication during the period that the bulk of Base Sample 1 covers.

To evaluate the results, we run our stacked DiD estimator on the structured quasi-treated and control samples generated for H1. Table 2.3 shows the stacked DiD estimates. As we expected, we observe statistically non-significant changes across the board for all of the indicators, providing support for hypothesis 1. A question that might arise at this stage is, could endogenous selection not be at play here? We defer the answer to this question to the discussion section, where we analyze the collective implications of all of the hypotheses test results. In the meantime, we note that the results are consistent with our hypothesis.

2.4.2. What happens when the numerical strength of the female directors is increased?

Next, we address the second question: what happens to the five firm-level risk-taking proxies, the firm value proxy, and the volatility of the market-adjusted return when previously gender-skewed majority male boards with a male CEO and a female director add more ordinary female directors to the board in non-leading roles? Our ex-ante expectations derive from the first of the second part of our hypotheses (i.e., hypothesis 2a or H2a). We structured our treated and control samples and chose the event windows to enhance the chances that non-disruptive board conditions would have been met, more or less. Thus, in this case, we expect statistically significant changes in the firm-level risk proxies. As Leary and Roberts (2005) point out, the dynamic pecking order theory predicts that firms are more concerned about excessively high leverage than excessively low leverage. The prediction is consistent with Wabara (2021a) and, indeed, hypothesis 2a. Hence, to the extent that the existing leverage is excessive, we would expect female directors to use their increasing influence on the board to call for a reduction of the excess.

We would also expect the changes in the other firm-level risk indicators to be consistent with Rajan and Zingales (1995). If leverage is reduced significantly, for example, we would typically expect a reduction in the volatility of the market-adjusted return. However, given the period that much of Base Sample 1 covers, this reduction may not necessarily be statistically significant. Again, to evaluate the results, we run our stacked DiD estimator on the structured quasi-treated and control samples generated for H2a.

Table 2.4 shows the stacked DiD estimates. We find a reduction in leverage that is statistically significant at the 1% level. Consistent with Rajan and Zingales (1995), this reduction in leverage is accompanied by increases in profitability and Tobin's q that are also statistically significant at the 1% level. We also observe a reduction in tangibility that is, however, not statistically significant. Nevertheless, we observe an increase in the cash/asset ratio that is statistically significant at the 5% level. We also observe a decrease in the volatility of the cash/asset ratio that is statistically significant at the 1% level. As expected for these outcomes, we observe a small reduction in the volatility of the market-adjusted return that is, however, not statistically significant.

Overall, and to the extent that increases in profitability and Tobin's q reflect improvements in firm performance, it would seem that the leverage reduced might have been the excessive component, such that the performance of the firm improves as a result. The latter is consistent with our hypothesis that with increasing power and influence, female directors will reduce the excessive risk-taking behavior of the firm.

The above notwithstanding, two questions still need to be answered. First, could these results have been driven by external equity fundraising events (e.g., see Baker and Wurgler, 2002)? Second, what roles might other endogenous selection motives have played to generate these results? Again, we defer the answer to the latter question to the discussion section, where we analyze the collective implications of all of the hypotheses test results. In the meantime, we note that the results are consistent with our hypothesis. However, we deal with the former question (about equity market timing) just below.

2.4.2.1. Does equity market timing drive the results?

Suppose some equity market timing events drive the above-described results for hypothesis 2a. In that case, according to Baker and Wurgler (2002), the results should also be accompanied by an increase in net equity issue and not necessarily an increase in retained earnings. Hence, we run our stacked DiD estimator on the structured quasi-treated and control samples generated for H2a. This time, we do so for variables like net equity issues, newly retained attained earnings, and asset growth as well.

Table 2.5 shows the stacked difference-in-differences regression results on the structured quasi-experimental treated and control samples corresponding to H2a. Instead of a statistically significant increase, we find a statistically non-significant reduction in net equity issues. We also see a statistically non-significant reduction in net debt issues and a statistically non-significant uptick in total asset growth. However, we find an increase in newly retained earnings statistically significant at the 5% level.

Overall, we note that our results are consistent with Rajan and Zingales (1995) but not compatible with Baker and Wurgler (2002), suggesting that the impact of the increasing power/influence of the female directors (via greater numerical strength) on leverage and its determinants are not coincident with some equity market timing events.

2.4.2.2. What then drives the H2a test results?

In effect, through what channel are the female directors potentially influencing the managers to generate the H2a test results? Stonehill (1975) documents that a central goal of financial executives in many countries appears to be guaranteeing the financial stability of their company and the availability of funds. This sentiment might explain the increase in retained earnings, especially if the more influential female directors emphasize such concerns to the managers. Strebulaev (2007) also suggests that firms might sell assets to pay down debt to help nip future liquidity crises in the bud. If the latter is the case, then it may be that the assets sold are the less productive ones since firm value and profitability improve, and we do not observe a statistically significant change in total asset growth.

In any case, to evaluate the latter, we estimate stacked DiDs on physical asset sales and asset turnover. We show the results in Table 2.6. Consistent with the preceding conjectures, we find an increase in physical asset sales significant at the 5% level and an increase in asset turnover that is significant at the 10% level. These are in addition to the statistically significant increases in cash/asset and newly retained earnings previously reported. In general, the sum of our results suggests an activated managerial action to minimize the risk of a future liquidity crisis and improve profitability and firm value.

Next, we switch focus to the non-token aggregate position of the female directors.

2.4.3. What happens when the aggregate power of the female directors is increased?

What then happens to the five firm-level risk-taking proxies, the firm value proxy, and the volatility of the market-adjusted return when the power and influence of the female subgroup of directors are increased (via non-token aggregate position)? We use two sets of structured quasi-experimental treated and control samples corresponding to hypotheses 2b(1) and 2b(2), respectively, to address this question. For both hypotheses 2b(1) and 2b(2), we use the same treatment events involving male-to-female CEO transitions since the position of the CEO is a highly influential one both on the board and in the firm in general, especially so for publicly listed US firms. Typically, CEO transitions generate a substantial amount of news and communication with the market. Hence, we expect any changes to the volatility of the market return to be statistically significant. Also, for hypotheses 2b(1) and 2b(2), we expect the changes, if any, to be a reduction, consistent with the idea that influential female directors will reduce the excessive risk of the firm.

The main difference between hypotheses 2b(1) and 2b(2) is as follows. Hypothesis 2b(1) uses a control subsample involving male-to-male CEO transitions in firms with homogenous male boards. However, hypothesis 2b(2) uses a control subsample involving male-to-male CEO transitions in firms with gender-skewed majority male boards with one or more female directors. This difference has a vital implication for the results of the hypothesis tests. Suppose the power and influence component of our hypotheses is proper. In that case, the examination of hypothesis 2b(1) should also reveal some firm-level effects consistent with reducing excessive risk. The test of hypothesis 2b(2) need not. The latter is because, given the results from the test of hypothesis 2a, the female directors in the control sample for hypothesis 2b(2) are likely to have a similar reduction effect on excessive risk to that of the female directors in the treated sample. Consequently, the difference-in-differences between the treated and control samples will most likely be non-significant.

We show the results of both tests in Table 2.7. Panels A and B show the test results for H2b(1) and H2b(2), respectively. First, both panels show reductions in the volatility of the market-adjusted return that are significant at the 5% level. These results are consistent with the conjectures that (i) influential female directors will reduce the excessive risk-taking of the firm and (ii) the event of CEO transition will generate substantial news and market communication to elicit a significant reaction from the market. Second, as expected, we find results that suggest a firm-level risk-reduction effect in the test results for H2b(1) but not for H2b(2), consistent with our power and influence hypothesis. Third, we also observe that the firm-level effects observed for H2b(1) are qualitatively compatible with the stylized facts in Rajan and Zingales (1995). Specifically, we find for H2b(1) a statistically non-significant increase in leverage accompanied by similarly statistically non-significant decreases in profitability and firm value. The tangibility increase is directionally consistent and is statistically significant at the 5% level.

Moreover, in untabulated results: we find that the net increase in the firm tangibility is associated with an increase in asset turnover, significant at the 5% level. We also find no statistically significant association between these results and external fundraising. Overall, these results, including a net increase in tangibility and asset turnover and the statistically non-significant reductions in profitability and firm value, are consistent with the notion of “excessive risk” reduction (e.g., see Aggarwal et al., 2010).

2.5. Robustness Analysis

2.5.1. Do our results change qualitatively with firms in the utilities and financial industry?

In the corporate governance literature, it is standard practice to exclude firms in the utilities and financial industries because of significant differences in regulatory oversight that can limit the board's influence on firm-level policies. Accordingly, we have followed this convention up to this point in our analyses. We also imposed an additional restriction on the treated and control samples. Specifically, we required that a firm stays in the treated (control) sample if there is at least one firm in the same Fama-French 48 industry in the control (treated) sample. However, the functional form that we adopt for our stacked DiD regression estimator is saturated with several fixed effects, including interactive industry-year fixed effects, to help deal with layers of heterogeneities across multiple dimensions (e.g., year, cohort, firm, industry by year). If these fixed effects do much of what we intend for them to do, then relaxing the convention above and our requirement for the structured samples should not change our results qualitatively.

Thus, we drop the convention and the requirement and add all previously excluded samples to verify the latter. We replicate all of our results up to this point using the enlarged samples and collect the new results as similarly named Tables in Appendix 2.2. We find that our previous results remain qualitatively unchanged, strengthening the support for our hypotheses and increasing our confidence in our empirical methodology.

2.5.2. Do the vertical stacks of the treated/control samples look meaningful?

The simulation results in the Internet Appendix set the stage for us to answer the question posed in this section comfortably.[45] We focus on the vertical stacks of the proxies in tests where the impacts of our board gender recomposition events are statistically significant. For example, we do so for hypothesis 2a focusing on the vertical stacks for leverage and Tobin's q. We also plot their normalized vertical stacks because the separation of the vertical stacks in the pre-event window suggests only tiny (if not statistically zero) differences in firm fixed effects. The dynamic distortions also appear little and consistent.

Figure 2.11 (A1) shows the vertical stack for the treated/control samples for leverage. We observe only minor distortions due to the presence of time fixed effects. This observation suggests the time fixed effects are not severe. Hence, the normalization of the vertical stacks would nicely mimic the image of the underlying data over the event window, W. Also, the separation between the treated/control samples in the pre-event window is minor, suggesting that, on average, the treated and control firms have firm fixed effects that are not far apart. Next, Figure 2.11 (A2) shows the normalization of the vertical stacks

[45] A quick summary of the insights are also available in section 2.3.4

for leverage, from which the image of the DiD can be quicly inferred. Overall, the vertical stacks and their normalizations look meaningful. The DiD is also visually consistent with the value of the DiD computed using the functional regression model.

Similarly, Figure 2.11 (B1) shows the vertical stack for the treated/control samples for Tobin's q. Again, we observe only minor distortions due to time fixed effects suggesting that the normalization of the vertical stacks would also nicely mimic the image of the underlying data over the entire event window, W. Also, the separation between the treated/control samples in the pre-event window is essentially zero, suggesting that, on average, the treated and control firms have about the same firm fixed effects. Accordingly, Figure 2.11 (B2) shows the normalization of the vertical stacks for Tobin's q, from which the image of the DiD can also be inferred. Overall, the vertical stacks and their normalizations look meaningful. The DiD is also visually consistent with the value of the DiD computed using the functional regression model. We, below, discuss the implications in detail.

2.6. Discussion

2.6.1. Do female directors simplistically reduce risk, or do they minimize excessive risk?

The ensemble of Figure 2.11 provides valuable insights into the most likely answer to the vital question of whether influential female directors simplistically reduce risk or primarily reduce excessive risk if they exist. We focus on both figures to draw insights.

We begin with leverage. Pre-event, both the treated and the control firms each had a male CEO, just one ordinary female director on the board, and (by normalization) precisely the same level of leverage. On the one hand, on the event date, more ordinary female directors join the treated boards, increasing the numerical strength of the female subgroup of directors. Subsequently, normalized leverage hardly changes. On the other hand, on the same event date, more ordinary male directors join the control boards, further increasing the numerical strength of the male subgroup of directors. Accordingly, the normalized leverage skyrockets and flattens out at a relatively much higher level.

Next, we look at Tobin's q. Again, pre-event, both the treated and the control firms each had a male CEO, just one ordinary female director on the board, and (by normalization) precisely the same level of Tobin's q. On the one hand, on the event date, more ordinary female directors join the treated boards, increasing the numerical strength of the female subgroup of directors. Subsequently, the normalized Tobin's q skyrockets and flattens out at a much higher level. On the other hand, on the same event date, more ordinary male directors join the control boards, further increasing the numerical strength of the male subgroup of directors. Accordingly, the normalized Tobin's q also increases but flattens out at a much lower level relative to that of the treated firms.

The combined picture from both normalized graphs suggests that both boards may have had close to optimal leverage in the pre-event window. However, in the post-event window, two contrasting things happen. On the one hand, the addition of more female directors coincides with the treated firms remaining at the optimal level of leverage and firm value rising to a commensurately higher level. On the other hand, the addition of more male directors in the counterfactual situation coincides with the uptake of additional leverage to a level significantly above the optimal level and firm value rising but not to a level commensurate with the optimal. Altogether, the preceding supports the notion that, with increasing power and influence on the board, the female directors counterfactually reduce (or prevent the uptake of) excessive leverage, consistent with our hypotheses.

We also arrive at the same conclusion by looking at the volatility of cash/asset, etc.

2.6.2. Is it power and influence or ex-ante endogenous director selection beliefs/motives?

Our structuring of (i) H2a in relation to H1, and (ii) H2b(1) in relation to H2b(2), and all the results discussed thus far help rule out the possibility that endogenous selection based on compatibility of risk preferences drives our findings. For example, suppose boards mainly choose directors based on the ex-ante compatibility of risk preferences between the incumbents and the new directors. In that case, by H2a, female directors have a relatively lower appetite for excessive risk; and the treated boards are similar in that they also have a relatively lower appetite for excessive risk. Conversely, male directors have a relatively higher appetite for excessive risk; and the control boards are similar in that they also have a relatively higher appetite for excessive risk. In sum, the gender of new directors mainly points to the relative differences in the appetites of the incumbents for excessive risk. However, applying the latter conclusion to H1, we should find a difference in risk-taking between the treatment and control boards. We do not, leading to a contradiction.[46]

The next question might relate to whether our results could not be explained by an endogenous selection of female directors based on ability[47] story (e.g., see Bertrand and Schoar, 2003). The answer is short. The choice of female directors based on ability is not a threat to or incompatible with our power and influence framework. For instance, suppose boards choose directors based on their ability to deal with excessive risk-taking one way or the other. Then by H2a, boards choose more female (male) directors to counterfactually reduce (increase) excessive risk. However, considering H1 and H2a together, the female directors still need greater numerical strength to get the job done. Hence, ex-ante motive alone is insufficient as female directors still need power/influence to get the job done.

[46] Moreover, the ex-ante motive of choosing directors based on the compatibility of risk preferences concedes that female directors have a relatively lower appetite for excessive risk-taking. These analyses transfer to H2b.
[47] Ability includes things like financial expertise, et cetera.

Overall, while our results establish power and influence as essential factors for the manifestation of the value-maximizing impact of the female subgroup of directors, we do not claim that it is impossible for a more welcoming "tone at the top" (e.g., see Bilings, Klein, and Shi, 2021) to also play a role in facilitating a board atmosphere in which female directors can have an impact. Our key point is that, even in that case, the female subgroup of directors must still be capable of exercising their influence (one way or another, soft or hard) for their value-adding preferences to manifest in the collective board decisions.

2.6.3. What are some of the implications of our results?

Our results confirm vital stylized facts about the determinants of leverage (e.g., Rajan and Zingales, 1995). We also document that increasing power and influence for the female directors might help firms operate at optimal leverage levels since female directors do not just reduce risk; they reduce excessive risk. Overall, our results have public discourse and policy implications. Precisely, they suggest that it would be more meaningful to begin shifting the academic and public discourse from whether diversity (board gender diversity, in particular) adds value to how best might the gender-diversification processes occur to ensure that the value that is inherent in diversity can be profitably unlocked.

2.7. Conclusion

Legal mandates and other forms of requirements for gender diversity on corporate boards continue to expand across geographies. In the USA, support for gender diversity on corporate boards is no longer a mere matter of principle. Yet, the business case for gender diversity on corporate boards remains inconclusive. The latter is underscored by the mixed extant empirical evidence in the academic literature on the impact of gender diversity on firms' financial performance. We argue that the mixed nature of the extant empirical evidence derives largely from the differential roles of board structure in the modulation of the impact of gender diversity. We pinpoint the differential roles of individual power and influence asymmetries within-board.

Broadly, we hypothesize that if a female director is unlikely to have any personal power or influence on the board, her addition to the board will have no significant impact on firm risk-taking and financial performance. However, with increasing power/influence on the board (via greater numerical strength or non-token aggregate position), female directors will tend to reduce the excessive risk-taking behavior of the firm and, to the extent that the gender-diversification is non-disruptive, this effect can feature significant increases in profitability and firm value.

Subsequently, we decompose our hypotheses and test them using sets of quasi-temporal event counterfactual samples drawn from board and financial data for publicly listed US firms. We use the stacked regression estimator to implement the tests and find results that are consistent with our hypotheses. Our results suggest that increasing power/influence for female directors might help firms operate at optimal leverage levels since female directors do not just reduce risk; they reduce excessive risk.

Our results also suggest that it might be more economical to begin moving the public (and academic) discussions and debates from whether diversity, board gender diversity, in particular, adds value to how best might the gender-diversification processes occur to ensure that the value that is inherent in diversity can be profitably unlocked.

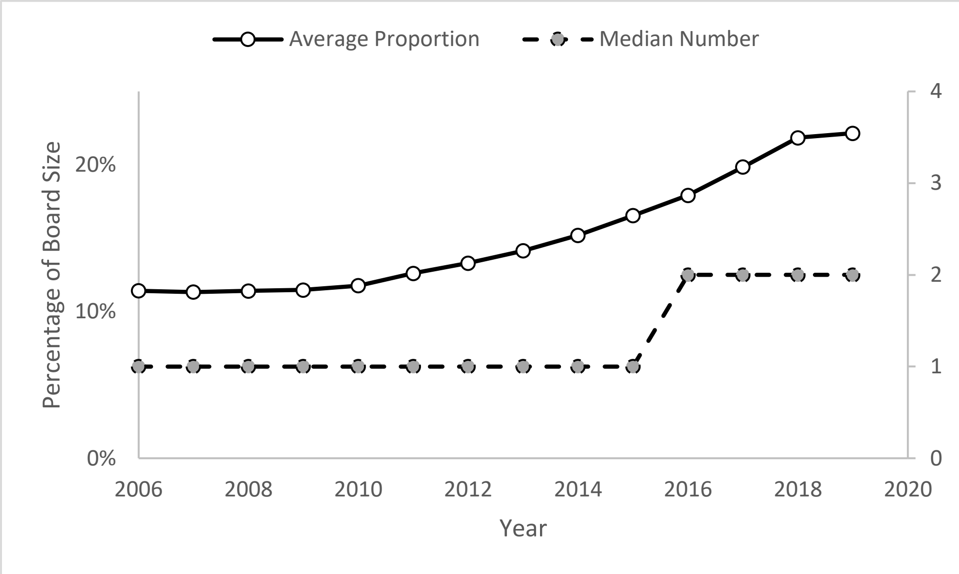
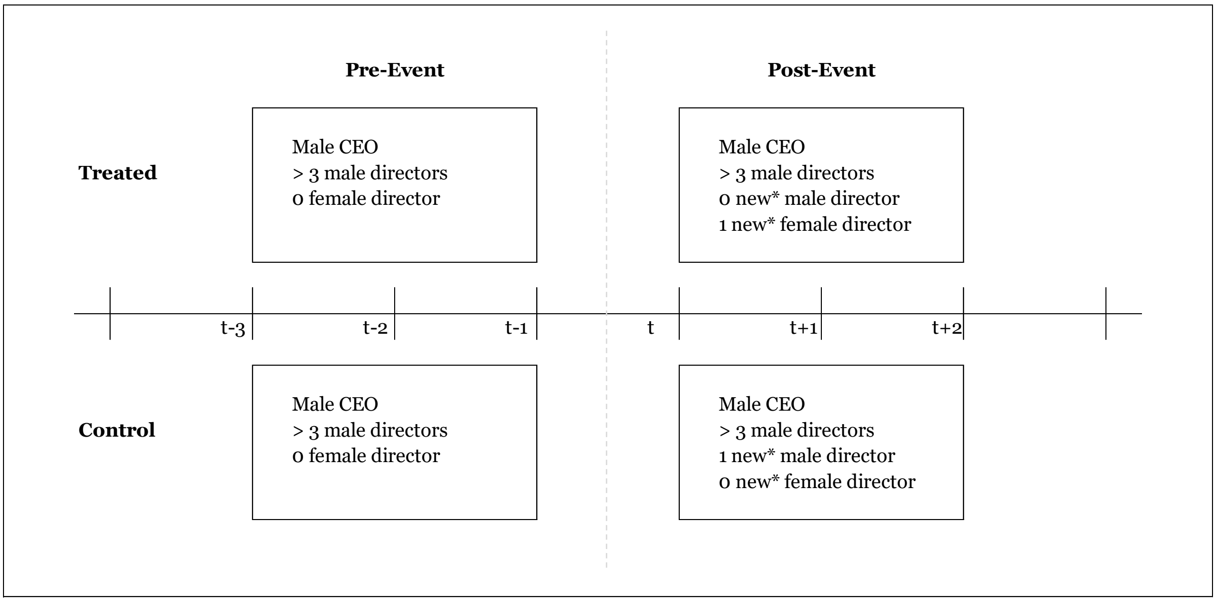


Figure 2.1
Annual board gender composition of publicly listed US firms
This figure shows the average proportion and the median number (both per year) of female directors on publicly listed US firms. It appears that following the global financial crisis in 2008, the average ratio of female directors on US boards began to rise steadily from 10% and now seems to be peaking at about 22%. The median number of female directors per board per year did not shift from one to two until 2016.



** New director stays in a non-leading role and no other major changes to board composition for at least 2 years*

Figure 2.2
Structuring the treated and control samples for hypothesis 1 (H1)

This hypothesis test seeks to evaluate the effect of adding one ordinary female director (i.e., a female director in a non-leading role) to a previously homogenous male board. In this case, the boards are homogenously male with male CEOs for at least three years before introducing the new ordinary female director (for the treated firms) or the new ordinary male director (for the control firms). We call this the period from t-3 to t-1, or in other sections, periods 1, 2, and 3. The first year in which the new director appears in the sample as a board member is the first of the three years post-event. This characterization is reasonable because new directors typically become part of the board before the end of the fiscal year that they first appear in the sample as board members. We call this period t, or in other sections, period 4. The board size only increases by one (i.e., the new director). The new directors stay for at least another two years, during which the board size/composition remains unchanged. We call this the period from t+1 to t+2, or in other sections, periods 5 and 6. The male CEOs are unchanged the three years before and the three years after. A firm stays in the treated (control) sample if there is at least one firm in the same Fama-French 48 industry in the control (treated) sample. To verify that our results remain qualitatively unchanged, we drop this condition in the robustness test.

	Base sample	*H1*	
Year	Number of observations	Control	Treated
(1)	*(2)*	*(3)*	*(4)*
2006	293	1.0%	1.7%
2007	524	3.1%	3.1%
2008	585	3.9%	4.6%
2009	611	4.9%	7.7%
2010	647	5.9%	10.0%
2011	682	6.5%	11.4%
2012	701	6.6%	11.7%
2013	729	5.2%	10.8%
2014	739	4.3%	10.1%
2015	728	3.7%	7.3%
2016	737	2.6%	4.3%
2017	736	1.2%	3.0%
2018	690	0.7%	2.2%
2019	261	1.1%	4.2%
All	8,663	3.8%	7.0%

Figure 2.3
Annual distribution of the treated and control samples for hypothesis 1
This Figure shows the yearly distribution of the treated and control samples for hypothesis test 1 (H1). Column 1 shows the fiscal year ends represented in Base Sample 1. Column 2 shows the number of firm-year observations per fiscal year in the base sample. Columns 3 and 4 show the results of the automated treated and control samples search. Specifically, Column 3 shows the proportion of the firm years that are in the control sample. On average, about 3.8% of the 8,663 observations are in the control sample. The majority of the control sample falls in the period between 2008 and 2016. Similarly, column 4 shows the proportion of the firm years that are in the treated sample. On average, about 7% of the 8,663 observations are in the treated sample. The majority of the treated sample also falls in the period between 2008 and 2016. Overall, the results suggest that following the 2008 financial crisis, previously homogenous male boards with male CEOs added at least one new director to the board in non-leading roles. However, of those new directors added, nearly twice as many were female directors. This interpretation is also consistent with the trend shown earlier in Figure 2.1. Notably, most of these endogenous board re-composition events happened before the #Me Too Movement (in 2017), the “Big Three” push (in 2017) for more women on the board, and the California gender quota law (in 2018).

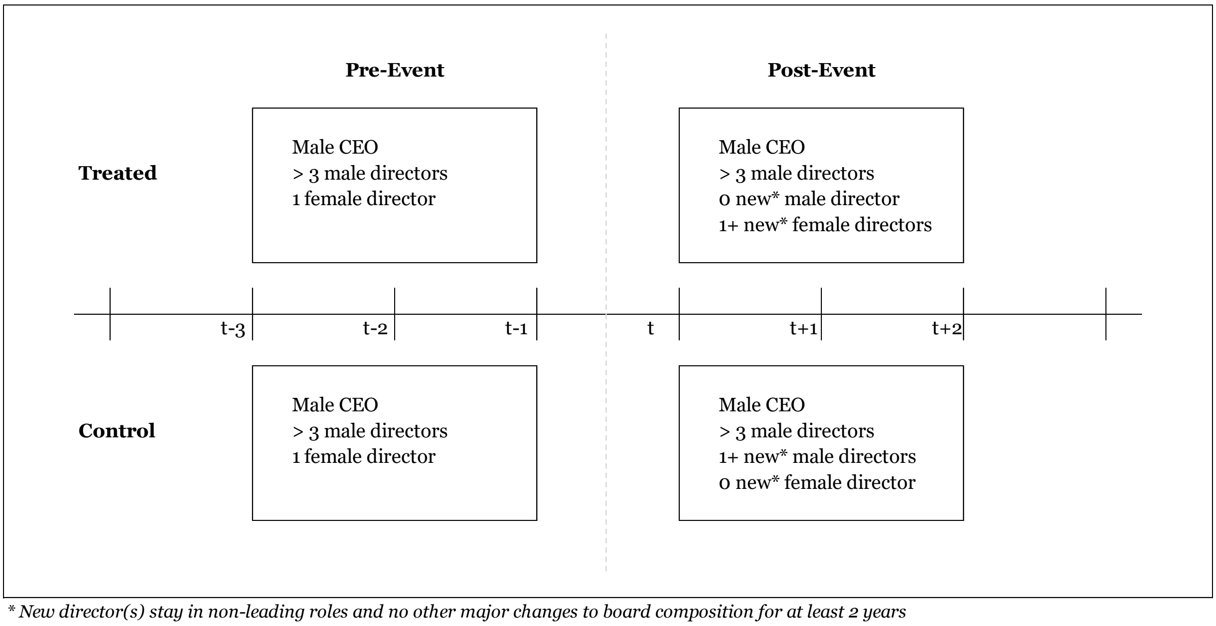
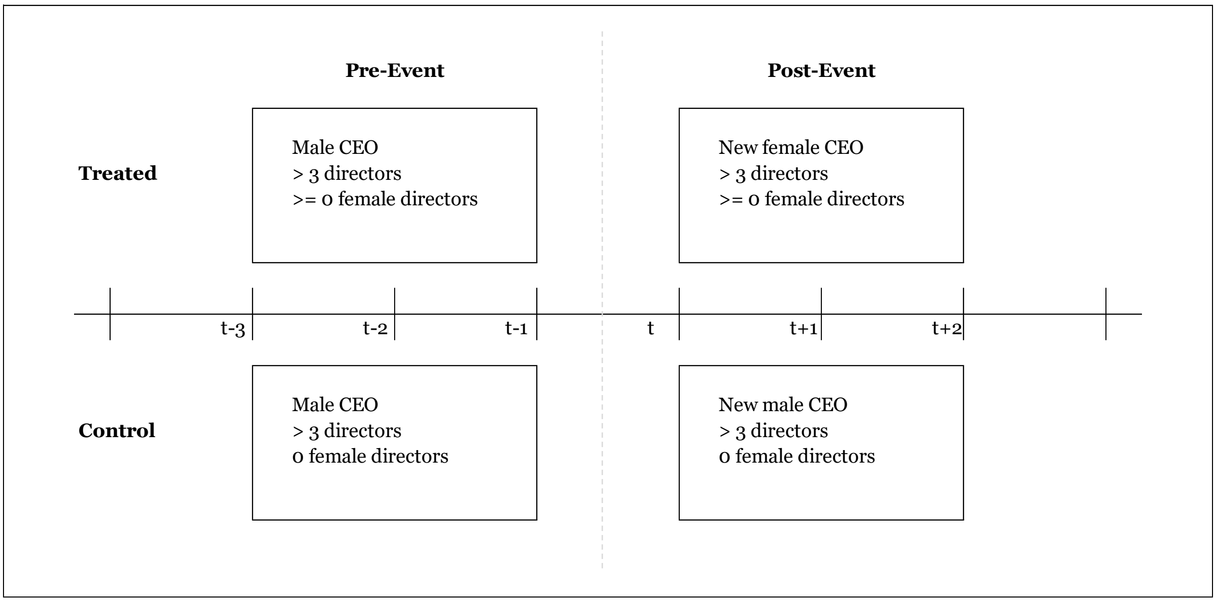


Figure 2.4
Structuring the treated and control samples for hypothesis 2a (H2a)
This hypothesis test evaluates the effect of adding one or more ordinary female directors (i.e., female directors in non-leading roles) to a previously majority male board with only one female director. In this case, the boards are skewed majority male with male CEOs and only one female director for at least three years before introducing an additional one or more ordinary female directors (for the treated firms) or ordinary male directors (for the control firms). Again, we call this the period from t-3 to t-1, or in other sections, periods 1, 2, and 3. The first year in which the new (additional one or more) ordinary directors appear in the sample as board members is the first of the three years post-event. Again, this characterization is reasonable because new directors typically become part of the board before the end of the fiscal year that they first appear in the sample as board members. We also call this period t, or in other sections, period 4. The board size only increases by the number of new directors added. The new directors stay for at least another two years, during which the board size/composition remains unchanged. We also call this the period from t+1 to t+2, or in other sections, periods 5 and 6. The male CEOs are unchanged the three years before and the three years after. A firm stays in the treated (control) sample if there is at least one firm in the same Fama-French 48 industry in the control (treated) sample. To verify that our results remain qualitatively unchanged, we drop this condition in the robustness test.

	Base sample	*H2a*	
Year	Number of observations	Control	Treated
(1)	*(2)*	*(3)*	*(4)*
2006	293	2.0%	1.0%
2007	524	2.9%	1.3%
2008	585	4.3%	2.4%
2009	611	5.6%	4.3%
2010	647	6.3%	5.6%
2011	682	7.9%	7.0%
2012	701	7.3%	8.7%
2013	729	6.9%	9.2%
2014	739	5.5%	8.5%
2015	728	4.0%	7.0%
2016	737	3.0%	5.7%
2017	736	1.5%	3.9%
2018	690	0.6%	2.3%
2019	261	1.1%	1.1%
All	8,663	4.5%	5.4%

Figure 2.5
Annual distribution of the treated and control samples for hypothesis 2a
This Figure shows the yearly distribution of the treated and control samples for hypothesis test 2a (H2a). Column 1 shows the fiscal year ends represented in Base Sample 1. Column 2 shows the number of firm-year observations per fiscal year in the base sample. Columns 3 and 4 show the results of the automated treated and control samples search. Specifically, Column 3 shows the proportion of the firm years that are in the control sample. On average, about 4.5% of the 8,663 observations are in the control sample. The majority of the control sample falls in the period between 2007 and 2016. Similarly, column 4 shows the proportion of the firm years that are in the treated sample. On average, about 5.4% of the 8,663 observations are in the treated sample. The majority of the treated sample falls in the period between 2009 and 2016. Overall, the results suggest that following the 2008 financial crisis, the previously skewed majority male boards with male CEOs and only one female director began to add at least one new director to the board in non-leading roles. However, of those new ordinary directors added by such boards, a majority of them were female directors. This interpretation is also consistent with the trend shown earlier in Figure 2.1. Notably also, the bulk of these endogenous board re-composition events happened before the #Me Too Movement (in 2017), before the "Big Three" push (in 2017) for more women on the board, and before the California gender quota law (in 2018).



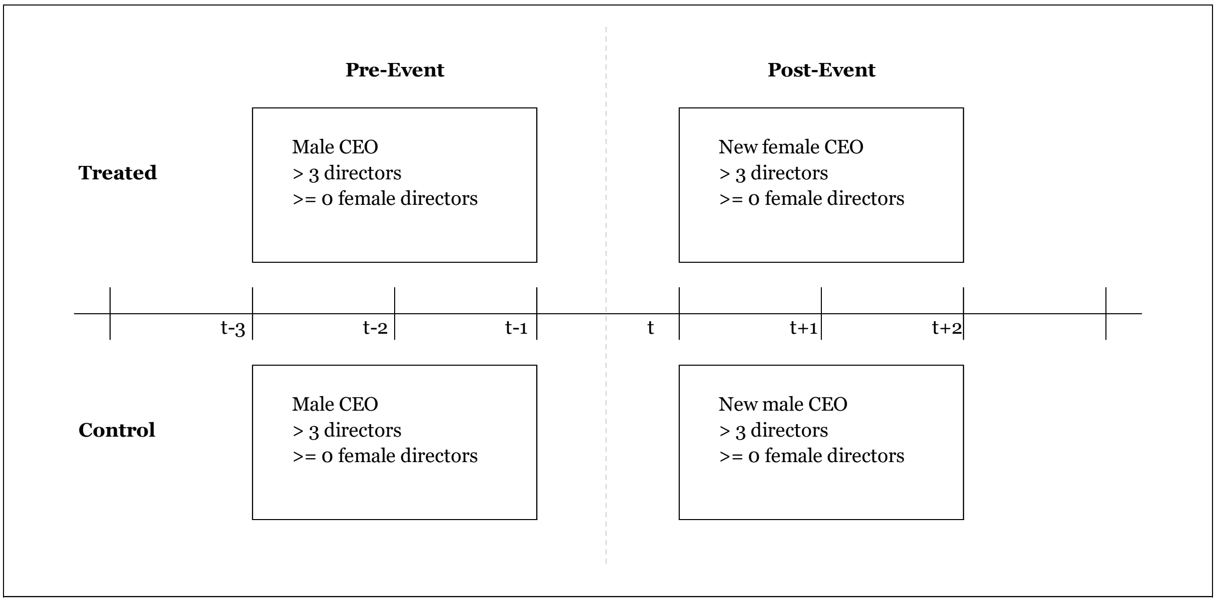
** New CEO stays for at least 2 years*

Figure 2.6
Structuring the treated and control samples for hypothesis 2b(1) (H2b(1))

This hypothesis test evaluates the effect of male-to-female CEO transition when the firms in the control sample do not have a female director on the board. In this case, the treated board has a unique male (i.e., not a joint) CEO and may have one or more female directors for at least three years before transitioning from the male CEO to a female CEO. The control board also has a unique male (i.e., not a joint) CEO but no female director for at least three years before transitioning from the male CEO to a new male CEO. We call this the period from t-3 to t-1, or in other sections, periods 1, 2, and 3. The first year in which the CEO transitions occur in the sample is the first of the three years post-event. Again, this characterization is reasonable because new directors typically become part of the board before the end of the fiscal year that they first appear in the sample as board members. We also call this period t, or in other sections, period 4. The board size does not increase by more than one (to accommodate the event that the new CEO was an outsider). The new CEO stays for at least another two years, during which the new board size/composition remains unchanged. We also call this the period from t+1 to t+2, or in other sections, periods 5 and 6. A firm stays in the treated (control) sample if there is at least one firm in the same Fama-French 48 industry in the control (treated) sample. To verify that our results remain qualitatively unchanged, we drop this condition in the robustness test.

	Base sample	H2b(1)	
Year	Number of observations	Control	Treated
(1)	(2)	(3)	(4)
2003	650	11.1%	1.4%
2004	674	14.1%	2.2%
2005	702	16.7%	2.8%
2006	751	18.6%	3.3%
2007	924	16.8%	3.8%
2008	905	15.8%	4.0%
2009	923	15.1%	5.1%
2010	939	15.0%	4.7%
2011	954	13.8%	4.7%
2012	961	13.1%	4.2%
2013	988	12.8%	3.9%
2014	1005	11.6%	4.0%
2015	1009	9.5%	2.7%
2016	1018	6.9%	2.0%
2017	1031	4.6%	1.5%
2018	1028	2.6%	1.4%
2019	769	2.3%	1.0%
All	15,231	11.6%	3.1%

Figure 2.7
Annual distribution of the treated and control samples for hypothesis 2b(1)
This Figure shows the yearly distribution of the treated and control samples for hypothesis test 2b(1) (H2b(1)). Column 1 shows the fiscal year ends represented in Base Sample 2. We restrict the firm years for Base Sample 2 to 2003-2019 (to remain broadly comparable to Base Sample 1, in which the fiscal firm years range from 2006 to 2019). Column 2 shows the number of firm-year observations per fiscal year in the base sample. Columns 3 and 4 show the results of the automated treated and control samples search. Specifically, Column 3 shows the proportion of the firm years that are in the control sample. On average, about 11.6% of the 15,231 observations are in the control sample. The majority of the control sample falls between 2003 and 2016, with a peak in 2006. Similarly, column 4 shows the proportion of the firm years that are in the treated sample. On average, about 3.1% of the 15,231 observations are in the treated sample. The majority of the treated sample falls between 2005 and 2015, with a peak in 2009. Overall, the results suggest that following the 2008 financial crisis, the bulk of CEO transitions for publicly listed US firms have been male-to-male, but a switch in the gender of CEOs from male to female appears to have picked up considerably in 2009. Nevertheless, these US firms with homogenous male boards remain 4x as likely to replace a male CEO with a male CEO as a gender-skewed majority male board replacing a male CEO with a female CEO. Notably, the bulk of these CEO transition events happened before the #Me Too Movement (in 2017), the "Big Three" push (in 2017) for more women on the board, and the California gender quota law (in 2018).



** New CEO stays for at least 2 years*

Figure 2.8
Structuring the treated and control samples for hypothesis 2b(2) (H2b(2))

This hypothesis test evaluates the effect of male-to-female CEO transition when the firms in the control sample may have a female director on the board. In this case, the treated board has a unique male (i.e., not a joint) CEO and may have one or more female directors for at least three years before transitioning from the male CEO to a female CEO. The control board also has a unique male (i.e., not a joint) CEO and may also have one or more female directors for at least three years before transitioning from the male CEO to a new male CEO. We call this the period from t-3 to t-1, or in other sections, periods 1, 2, and 3. The first year in which the CEO transitions occur in the sample is the first of the three years post-event. Again, this characterization is reasonable because new directors typically become part of the board before the end of the fiscal year that they first appear in the sample as board members. We also call this period t, or in other sections, period 4. The board size does not increase by more than one (to accommodate the event that the new CEO was an outsider). The new CEO stays for at least another two years, during which the new board size/composition remains unchanged. We also call this the period from t+1 to t+2, or in other sections, periods 5 and 6. A firm stays in the treated (control) sample if there is at least one firm in the same Fama-French 48 industry in the control (treated) sample. To verify that our results remain qualitatively unchanged, we drop this condition in the robustness test.

	Base sample	H2b(2)	
Year	Number of observations	Control	Treated
(1)	(2)	(3)	(4)
2003	650	19.8%	1.4%
2004	674	26.9%	2.2%
2005	702	32.5%	2.8%
2006	751	35.6%	3.3%
2007	924	35.7%	3.8%
2008	905	34.4%	4.0%
2009	923	33.6%	5.1%
2010	939	33.0%	4.7%
2011	954	32.3%	4.7%
2012	961	31.9%	4.2%
2013	988	30.3%	3.9%
2014	1005	28.1%	4.0%
2015	1009	23.1%	2.7%
2016	1018	17.8%	2.0%
2017	1031	11.5%	1.5%
2018	1028	6.3%	1.4%
2019	769	5.7%	1.0%
All	15,231	25.6%	3.1%

Figure 2.9
Annual distribution of the treated and control samples for hypothesis 2b(2)
This Figure shows the yearly distribution of the treated and control samples for hypothesis test 2b(2) (H2b(2)). Column 1 shows the fiscal year ends represented in Base Sample 2. Column 2 shows the number of firm-year observations per fiscal year in the base sample. Columns 3 and 4 show the results of the automated treated and control samples search. Specifically, Column 3 shows the proportion of the firm years that are in the control sample. On average, about 25.6% of the 15,231 observations are in the control sample. The majority of the control sample falls between 2003 and 2016, with a peak in 2006. Similarly, column 4 shows the proportion of the firm years that are in the treated sample. On average, about 3.1% of the 15,231 observations are in the treated sample. The majority of the treated sample falls between 2005 and 2015, with a peak in 2009. Overall, the results suggest that following the 2008 financial crisis, the bulk of CEO transitions for publicly listed US firms have been male-to-male, but a switch in the gender of CEOs from male to female appears to have picked up considerably in 2009. Nevertheless, these US firms with gender-skewed majority male boards remain over 8x as likely to replace a male CEO with a male CEO as to replace a male CEO with a female CEO. Notably also, the bulk of these CEO transition events happened before the #Me Too Movement (in 2017), the "Big Three" push (in 2017) for more women on the board, and the California gender quota law (in 2018).

Fama-French Industry	*H1*		*H2a*		*H2b(1)*		*H2b(2)*	
	Control	Treated	Control	Treated	Control	Treated	Control	Treated
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*
2 Food			6.2%	4.1%	8.3%	9.2%	15.5%	9.2%
7 Fun					11.0%	6.3%	24.6%	6.3%
8 Books					5.4%	7.1%	25.9%	7.1%
9 Hshld					6.9%	1.8%	15.6%	1.8%
10 Clths	2.9%	2.9%			8.1%	2.0%	31.1%	2.0%
11 Hlth	3.5%	4.2%	4.2%	4.2%				
12 MedEq	5.5%	11.0%	3.7%	11.0%				
13 Drugs			2.0%	3.7%	13.0%	1.9%	30.1%	1.9%
14 Chems	2.2%	6.6%	12.8%	6.6%	14.8%	3.6%	34.3%	3.6%
17 BldMt	2.5%	12.0%	4.6%	7.5%	14.8%	3.1%	24.8%	3.1%
18 Cnstr	13.8%	15.6%	2.8%	1.8%				
19 Steel	6.2%	6.8%	6.2%	6.8%	5.0%	2.1%	14.6%	2.1%
21 Mach	7.8%	6.5%	3.9%	2.6%	15.6%	2.0%	25.7%	2.0%
22 ElcEq	4.8%	13.6%	4.8%	5.6%				
23 Autos	6.7%	6.7%	9.4%	6.7%				
30 Oil	7.4%	11.7%	3.0%	6.0%	19.7%	1.6%	32.7%	1.6%
31 Util	x	x	x	x	x	x	x	x
32 Telcm					9.3%	7.6%	29.7%	7.6%
33 PerSv					7.6%	1.8%	30.0%	1.8%
34 BusSv	4.6%	5.8%	5.8%	4.0%	8.9%	2.3%	29.6%	2.3%
35 Comps	2.0%	9.9%	6.0%	9.3%	9.5%	1.2%	24.3%	1.2%
36 Chips	9.5%	10.3%	3.7%	3.7%	18.6%	2.1%	30.9%	2.1%
37 LabEq	4.2%	8.4%	4.2%	6.3%	18.0%	1.4%	30.0%	1.4%
38 Paper	5.2%	5.2%	5.2%	5.2%	22.5%	2.8%	30.5%	2.8%
39 Boxes					17.9%	3.7%	22.4%	3.7%
40 Trans	5.1%	11.6%	1.7%	6.8%	13.9%	1.0%	21.7%	1.0%
41 Whlsl			4.2%	8.2%				
42 Rtail			1.8%	6.3%	9.1%	6.8%	26.1%	6.8%
43 Meals					5.9%	5.2%	23.8%	5.2%
44 Banks	x	x	x	x	x	x	x	x
45 Insur	x	x	x	x	x	x	x	x
46 RlEst								
47 Fin	x	x	x	x	x	x	x	x

The values are proportions of the number of obervations in the bases samples
Columns 1 to 4 are based on Base Sample 1 and columns 5 to 8 are based on Base Sample 2
x represents observations excluded in the initial analyses but added back in the robustness tests

Figure 2.10
Industry distribution of the treated and control samples

This Figure shows the industry distribution of the treated and control samples for all four hypotheses. The values are the proportions of the number of observations in the bases samples. Columns 1 to 4 are computed based on the observations in Base Sample 1, and columns 5 to 8 are calculated based on the observations in Base Sample 2. *x* represents observations excluded in the initial analyses but added back in the robustness tests. Overall, the results show another dimension that we consider in estimating the stacked difference-in-differences (DiD).

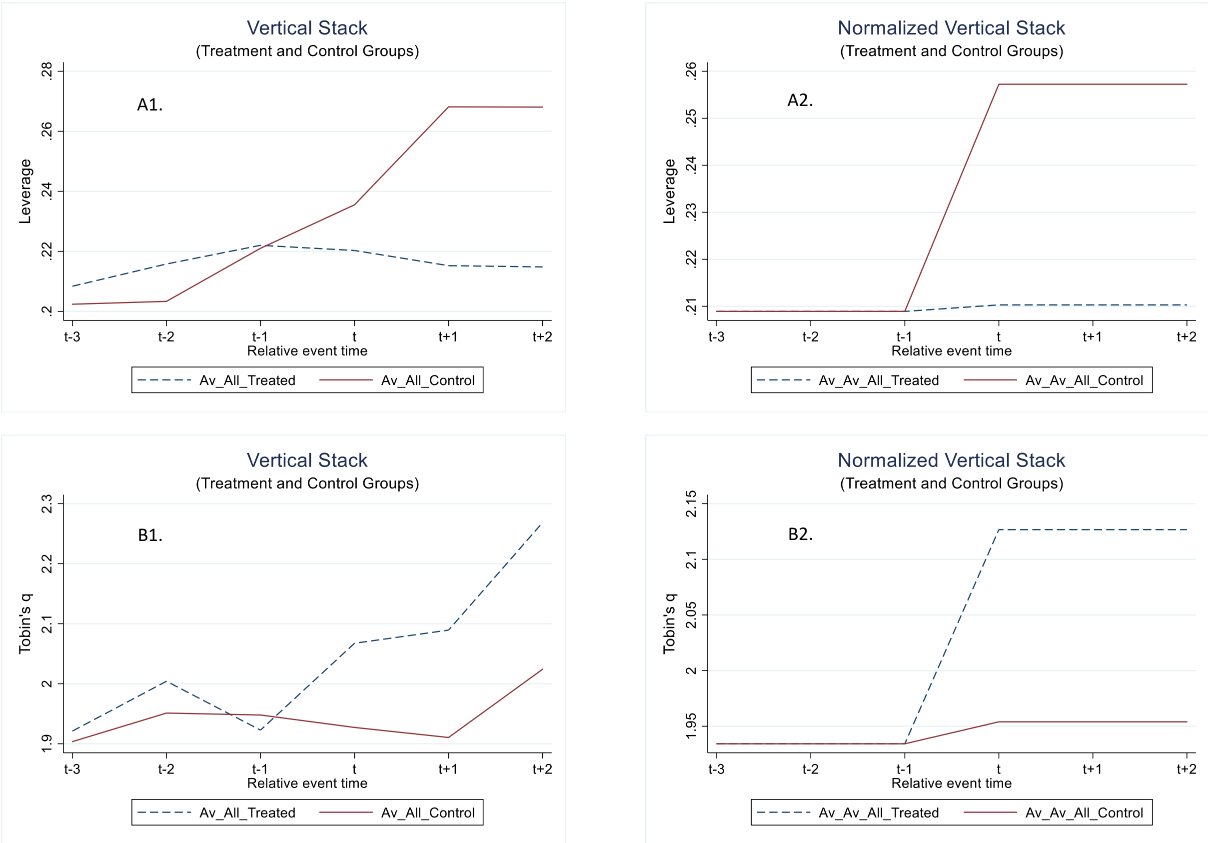


Figure 2.11
Vertical stacks and their normalizations (H2a: Leverage and Tobin's q)
This Figure plots the vertical stacks (A1 and B1) and their normalizations (A2 and B2) for leverage and Tobin's q, respectively. The images are generated using the quasi-temporal event counterfactual samples corresponding to hypothesis 2a (H2a). As described in section 2.3.4, the vertical stacks are computed using equations (1) and (2) as follows:

$$RI^{Stacked}_{\omega,\ Treated} = 1\{W = \omega,\ \ Treated = 1\} * \frac{1}{N_1}\sum_{i=1}^{N_1} RI_i \qquad (1)$$

$$RI^{Stacked}_{\omega,\ Control} = 1\{W = \omega,\ \ Treated = 0\} * \frac{1}{N_0}\sum_{i=1}^{N_0} RI_i \qquad (2)$$

Where RI is the risk indicator or proxy variable of interest. N_1 and N_0 are the number of units (firms, in the context of this study) in the treated and control groups, respectively. $W = \{t-3,\ t-2,\ t-1,\ t,\ t+1,\ t+2\}$. $Treated = \{0, 1\}$ where one (1) corresponds to a treated sample and zero (0) corresponds to a control sample. Drawing from the insights we developed in the Internet Appendix, these vertical stacks look well-behaved with non-severe time distortions and hardly any difference in the average firm fixed effects. As we also show in the Internet Appendix, when the distortions from the time fixed effects are not severe, the normalization of the vertical stacks nicely mimics the image of the underlying data over the event window, W. Normalization takes the half-window averages of the vertical stacks and reduces their pre-event differences to zero, such that the post-event difference between the treated and control gives a direct estimate of the stacked DiD. Inferring the DiD from the post-window difference is the goal.

Table 2.1. Base sample 1 (for H1 and H2a) – summary statistics

Leverage is the ratio of total debt to total assets. Tangibility is the ratio of PPE to total assets. Operating profitability is the ratio the operating income before depreciation to total assets. Tobin's q is total assets minus the book value of equity plus the market value of equity, divided by the book value of total assets. Total q is a contributed data from Peters and Taylor (2017), who define it as the aggregate market value scaled by total capital, the sum of the physical and intangible capital stocks. Log(Sales) is the natural logarithm of sales. Net equity issue is change in book equity minus change in balance sheet retained earnings, divided by total assets. Newly retained earnings is change in balance sheet retained earnings, divided by total assets. Asset growth is change in total assets, divided by lagged total assets. Net debt issue is change in total assets minus change in book equity, divided by total assets. Physical asset sales is sale of PPE, divided by lagged total assets. Asset turnover is sales divided by total assets. Cash/asset is the ratio of cash to total assets. Volatility of the Cash/asset is that standard deviation of Cash/asset computed over 8 preceding quarters. Volatility of market-adjusted return is standard deviation (computed over 36 preceding months) of the stock return less same period return on the CRSP value-weighted portfolio of NYSE/Amex/Nasdaq stocks. Board size is the number directors on the board. Female directors is the number of female directors on the board. Fraction of female directors is female directors, divided by board size. Independent directors is the number of independent directors on the board. Fraction of independent directors is independent directors, divided by board size. Female independent directors is number of female independent directors, divided by independent directors. Non-CEO leader is a leader on the board who is not the CEO (e.g., lead director). Fraction of female non-CEO leaders is the number of female non-CEO leaders, divided by non-CEO leaders. Female chairman is a dummy variable that takes the value if the chairman is female, and zero otherwise. Female CEO is a dummy variable that takes the value if the CEO is female, and zero otherwise. Female chairman/CEO is a dummy variable that takes the value if the chairman/CEO is female, and zero otherwise. Data source: Compustat, ISS Directors Data.

Variable	Number of observations	Mean	Standard deviation	Min	Median	Max
Firm characteristics						
Leverage	8,663	0.22	0.17	0.00	0.21	0.85
Tangibility	8,663	0.24	0.21	0.00	0.17	0.96
Operating profitability	8,663	0.14	0.08	-0.59	0.13	0.95
Tobin's q	8,063	2.01	1.24	0.45	1.66	14.65
Total q	6,698	1.27	1.43	-1.95	0.90	26.27
Log(Sales)	8,662	7.91	1.49	3.12	7.78	11.33
Net equity issue	7,935	0.00	0.08	-0.86	0.00	1.47
Newly retained earnings	8,663	0.02	0.10	-1.75	0.03	0.70
Asset growth	8,663	0.10	0.33	-0.73	0.05	12.19
Net debt issue	7,935	0.03	0.11	-1.51	0.02	0.75
Physical asset sales	6,154	0.00	0.01	0.00	0.00	0.31
Asset turnover	8,663	1.17	0.75	0.00	0.97	5.94
Cash / asset	8,663	0.11	0.11	0.00	0.08	0.72
Volatility of cash / asset	5,124	0.09	0.08	0.00	0.07	0.85
Volatility of market-adjusted return	8,634	0.08	0.03	0.02	0.08	0.25
Board characteristics						
Board size	8,663	9.26	2.05	3	9	22
Female directors	8,663	1.46	1.14	0	1	7
Fraction of female directors	8,663	0.15	0.11	0.00	0.14	0.75
Independent directors	8,663	6.80	2.77	0	7	20
Fraction of independent directors	8,663	0.73	0.23	0.00	0.80	1
Female independent directors	8,663	1.21	1.09	0	1	6
Fraction of female independent directors	8,024	0.17	0.13	0.00	0.17	1
Non-CEO leader	8,663	0.88	0.55	0	1	3
Fraction of female non-CEO leaders	6,753	0.06	0.24	0.00	0.00	1
Female chairman	8,663	0.03	0.16	0	0	1
Female CEO	8,663	0.03	0.18	0	0	1
Female chairman / CEO	8,663	0.01	0.11	0	0	1

Table 2.2. Base sample 2 (for H2b) – summary statistics

Leverage is the ratio of total debt to total assets. Tangibility is the ratio of PPE to total assets. Operating profitability is the ratio the operating income before depreciation to total assets. Tobin's q is total assets minus the book value of equity plus the market value of equity, divided by the book value of total assets. Total q is a contributed data from Peters and Taylor (2017), who define it as the aggregate market value scaled by total capital, the sum of the physical and intangible capital stocks. Log(Sales) is the natural logarithm of sales. Net equity issue is change in book equity minus change in balance sheet retained earnings, divided by total assets. Newly retained earnings is change in balance sheet retained earnings, divided by total assets. Asset growth is change in total assets, divided by lagged total assets. Net debt issue is change in total assets minus change in book equity, divided by total assets. Physical asset sales is sale of PPE, divided by lagged total assets. Asset turnover is sales divided by total assets. Cash/asset is the ratio of cash to total assets. Volatility of the Cash/asset is that standard deviation of Cash/asset computed over 8 preceding quarters. Volatility of market-adjusted return is standard deviation (computed over 36 preceding months) of the stock return less same period return on the CRSP value-weighted portfolio of NYSE/Amex/Nasdaq stocks.

Variable	Number of observations	Mean	Standard deviation	Min	Median	Max
Firm characteristics						
Leverage	15,231	0.21	0.17	0.00	0.20	0.87
Tangibility	15,231	0.25	0.22	0.00	0.18	0.96
Operating profitability	15,224	0.13	0.12	-2.67	0.13	0.95
Tobin's q	14,257	2.03	1.35	0.23	1.63	19.55
Total q	12,586	1.34	1.73	-4.44	0.89	35.90
Log(Sales)	15,201	7.54	1.69	-3.54	7.49	11.33
Net equity issue	14,034	-0.01	2.85	-337.42	0.01	2.19
Newly retained earnings	15,229	0.04	2.74	-5.22	0.03	337.85
Asset growth	15,231	0.12	0.36	-0.93	0.06	12.19
Net debt issue	14,035	0.03	0.22	-14.01	0.02	0.88
Physical asset sales	11,006	0.00	0.02	0.00	0.00	0.79
Asset turnover	15,231	1.20	0.79	0.00	1.01	5.94
Cash / asset	15,231	0.12	0.12	0.00	0.08	1.00
Volatility of cash / asset	7,397	0.09	0.09	0.00	0.07	0.96
Volatility of market-adjusted return	14,503	0.09	0.04	0.02	0.08	0.27

Table 2.3
Hypothesis test 1 (H1)
This table reports the stacked difference-in-differences regression results on the structured quasi-experimental treated and control samples corresponding to H1.

$$RI_{i,j,c,t} = \beta * Treated\_Post_{i,j,c,t} + p_t + m_c + \alpha_i + \vartheta_{j,t} + u_{i,j,c,t}$$

β captures the stacked difference-in-differences effect on each of the risk, firm value, or market reaction indicators ($RI_{i,j,c,t}$, named just above the column numbers, computed over the 3-year pre-event and 3-year post-event estimation windows for a firm i, in industry j, treated in year t, and belongs to cohort c). $Treated * Post$ is the interaction of the treated and post dummy variables and p_t, m_c, α_i, and $\vartheta_{j,t}$ are the year, cohort, firm, and industry-year fixed effects, respectively, that help minimize the heterogeneities across those dimensions. Leverage is the ratio of total debt to total assets. Tangibility is the ratio of PPE to total assets. Operating profitability is the ratio the operating income before depreciation to total assets. Cash/asset is the ratio of cash to total assets. Volatility of the Cash/asset is that standard deviation of Cash/asset computed over 8 preceding quarters. Tobin's q is total assets minus the book value of equity plus the market value of equity, divided by the book value of total assets. Volatility of market-adjusted return is standard deviation (computed over 36 preceding months) of the stock return less same period return on the CRSP value-weighted portfolio of NYSE/Amex/Nasdaq stocks.

	Risk indicators (within-firm)					*Firm value*	*Market's reaction*
	Leverage	Tangibility	Operating Profitability	Cash/Asset	Volatility of Cash/Asset	Tobin's q	Volatility of Market-Adjusted Return
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Treated * Post	-0.0060	-0.0038	-0.0036	0.0179	-0.0116	0.0271	0.0004
	(0.0120)	(0.0055)	(0.0077)	(0.0114)	(0.0179)	(0.0989)	(0.0026)
<u>Fixed effects</u>							
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry*Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.882	0.988	0.784	0.839	0.674	0.874	0.878
N	739	739	739	739	389	691	727

* p<0.1; ** p<0.05; *** p<0.01
Standard errors in parentheses

Table 2.4
Hypothesis test 2a (H2a)
This table reports the stacked difference-in-differences regression results on the structured quasi-experimental treated and control samples corresponding to H2a.

$$RI_{i,j,c,t} = \beta * Treated\_Post_{i,j,c,t} + p_t + m_c + \alpha_i + \vartheta_{j,t} + u_{i,j,c,t}$$

β captures the stacked difference-in-differences effect on each of the risk, firm value, or market reaction indicators ($RI_{i,j,c,t}$, named just above the column numbers, computed over the 3-year pre-event and 3-year post-event estimation windows for a firm i, in industry j, treated in year t, and belongs to cohort c). $Treated * Post$ is the interaction of the treated and post dummy variables and p_t, m_c, α_i, and $\vartheta_{j,t}$ are the year, cohort, firm, and industry-year fixed effects, respectively, that help minimize the heterogeneities across those dimensions. Leverage is the ratio of total debt to total assets. Tangibility is the ratio of PPE to total assets. Operating profitability is the ratio the operating income before depreciation to total assets. Cash/asset is the ratio of cash to total assets. Volatility of the Cash/asset is that standard deviation of Cash/asset computed over 8 preceding quarters. Tobin's q is total assets minus the book value of equity plus the market value of equity, divided by the book value of total assets. Volatility of market-adjusted return is standard deviation (computed over 36 preceding months) of the stock return less same period return on the CRSP value-weighted portfolio of NYSE/Amex/Nasdaq stocks.

	Risk indicators (within-firm)					*Firm value*	*Market's reaction*
	Leverage	Tangibility	Operating Profitability	Cash/Asset	Volatility of Cash/Asset	Tobin's q	Volatility of Market-Adjusted Return
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Treated * Post	-0.0508	-0.0029	0.0373	0.0179	-0.0447	0.2952	-0.0003
	(0.0109)***	(0.0073)	(0.0073)***	(0.0077)**	(0.0161)***	(0.0823)***	(0.0022)
Fixed effects							
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry*Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.920	0.982	0.840	0.910	0.708	0.908	0.909
N	634	634	634	634	376	563	621

* $p<0.1$; ** $p<0.05$; *** $p<0.01$
Standard errors in parentheses

Table 2.5
Are the H(2a) test results driven by equity market timing?

This table reports the stacked difference-in-differences regression results on the structured quasi-experimental treated and control samples corresponding to H2a. We add additional variables to evaluate whether the results in Table 2.4 are driven by equity market timing. The results are consistent with Rajan and Zingales (1995). However, they are not consistent with Baker and Wurgler (2002), suggesting that the effects of increasing power/influence of the female directors (via greater numerical strength) on leverage and its determinants are not coincident with equity market timing.

$$RI_{i,j,c,t} = \beta * Treated\_Post_{i,j,c,t} + p_t + m_c + \alpha_i + \vartheta_{j,t} + u_{i,j,c,t}$$

β captures the stacked difference-in-differences effect on each of the firm-level risk or equity market timing indicators ($RI_{i,j,c,t}$, named just above the column numbers, computed over the 3-year pre-event and 3-year post-event estimation windows for a firm i, in industry j, treated in year t, and belongs to cohort c). $Treated * Post$ is the interaction of the treated and post dummy variables and p_t, m_c, α_i, and $\vartheta_{j,t}$ are the year, cohort, firm, and industry-year fixed effects, respectively, that help minimize the heterogeneities across those dimensions. Leverage is the ratio of total debt to total assets. Tangibility is the ratio of PPE to total assets. Operating profitability is the ratio the operating income before depreciation to total assets. Tobin's q is total assets minus the book value of equity plus the market value of equity, divided by the book value of total assets. Net debt issue is change in total assets minus change in book equity, divided by total assets. Net equity issue is change in book equity minus change in balance sheet retained earnings, divided by total assets. Newly retained earnings is change in balance sheet retained earnings, divided by total assets. Asset growth is change in total assets, divided by lagged total assets.

	H2a: Consistent with Zingales (1995)				*H2a: Not consistent with Baker and Wurgler (2002)*			
	Leverage	Tangibility	Operating Profitability	Tobin's q	Net Debt Issue	Net Equity Issue	Newly Retained Earnings	Asset Growth
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*
Treated * Post	-0.0508	-0.0029	0.0373	0.2952	-0.0113	-0.0022	0.0362	0.0165
	(0.0109)***	(0.0073)	(0.0073)***	(0.0823)***	(0.0230)	(0.0148)	(0.0172)**	(0.0356)
Fixed effects								
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry*Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.920	0.982	0.840	0.908	0.515	0.583	0.508	0.503
N	634	634	634	563	460	460	529	529

* p<0.1; ** p<0.05; *** p<0.01
Standard errors in parentheses

Table 2.6
Does activated managerial action drive the H(2a) test results?
This table reports the stacked difference-in-differences regression results on the structured quasi-experimental treated and control samples corresponding to H2a. We add additional variables to evaluate whether activated managerial actions drive the results in Table 2.4.

$$RI_{i,j,c,t} = \beta * Treated\_Post_{i,j,c,t} + p_t + m_c + \alpha_i + \vartheta_{j,t} + u_{i,j,c,t}$$

β captures the stacked difference-in-differences effect on each of the firm-level risk or managerial action indicators ($RI_{i,j,c,t}$, named just above the column numbers, computed over the 3-year pre-event and 3-year post-event estimation windows for a firm i, in industry j, treated in year t, and belongs to cohort c). $Treated * Post$ is the interaction of the treated and post dummy variables and p_t, m_c, α_i, and $\vartheta_{j,t}$ are the year, cohort, firm, and industry-year fixed effects, respectively, that help minimize the heterogeneities across those dimensions. Leverage is the ratio of total debt to total assets. Tangibility is the ratio of PPE to total assets. Operating profitability is the ratio the operating income before depreciation to total assets. Tobin's q is total assets minus the book value of equity plus the market value of equity, divided by the book value of total assets. Physical asset sales is sale of PPE, divided by lagged total assets. Asset turnover is sales divided by total assets. Cash/asset is the ratio of cash to total assets. Newly retained earnings is change in balance sheet retained earnings, divided by total assets. Asset growth is change in total assets, divided by lagged total assets.

	H2a: Consistent with Zingales (1995)				*H2a: consistent with activated managerial action*			
	Leverage	Tangibility	Operating Profitability	Tobin's q	Physical Asset Sales	Asset Turnover	Cash/Asset	Newly Retained Earnings
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*
Treated * Post	-0.0508	-0.0029	0.0373	0.2952	0.0046	0.0838	0.0179	0.0362
	(0.0109)***	(0.0073)	(0.0073)***	(0.0823)***	(0.0020)**	(0.0438)*	(0.0077)**	(0.0172)**
Fixed effects								
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry*Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.920	0.982	0.840	0.908	0.938	0.973	0.910	0.508
N	634	634	634	563	377	529	634	529

* p<0.1; ** p<0.05; *** p<0.01
Standard errors in parentheses

Table 2.7
Hypothesis test 2b (H2b)
This table reports the stacked difference-in-differences regression results on the structured quasi-experimental treated and control samples corresponding to H2b. Panels A and B show the results for hypotheses 2b(1) and 2b(2), respectively. The main difference between hypotheses 2b(1) and 2b(2) is as follows. Hypothesis 2b(1) uses a control subsample involving male-to-male CEO transitions in firms with homogenous male boards. However, hypothesis 2b(2) uses a control subsample involving male-to-male CEO transitions in firms with gender-skewed majority male boards with one or more female directors.

$$RI_{i,j,c,t} = \beta * Treated\_Post_{i,j,c,t} + p_t + m_c + \alpha_i + \vartheta_{j,t} + u_{i,j,c,t}$$

β captures the stacked difference-in-differences effect on each of the risk, firm value, or market reaction indicators ($RI_{i,j,c,t}$, named just above the column numbers, computed over the 3-year pre-event and 3-year post-event estimation windows for a firm i, in industry j, treated in year t, and belongs to cohort c). $Treated * Post$ is the interaction of the treated and post dummy variables and p_t, m_c, α_i, and $\vartheta_{j,t}$ are the year, cohort, firm, and industry-year fixed effects, respectively, that help minimize the heterogeneities across those dimensions. Leverage is the ratio of total debt to total assets. Tangibility is the ratio of PPE to total assets. Operating profitability is the ratio the operating income before depreciation to total assets. Cash/asset is the ratio of cash to total assets. Volatility of the Cash/asset is that standard deviation of Cash/asset computed over 8 preceding quarters. Tobin's q is total assets minus the book value of equity plus the market value of equity, divided by the book value of total assets. Volatility of market-adjusted return is standard deviation (computed over 36 preceding months) of the stock return less same period return on the CRSP value-weighted portfolio of NYSE/Amex/Nasdaq stocks.

Panel A: H2b(1)	*Risk indicators (within-firm)*					*Firm value*	*Market's reaction*
	Leverage	Tangibility	Operating Profitability	Cash/Asset	Volatility of Cash/Asset	Tobin's q	Volatility of Market-Adjusted Return
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Treated * Post	0.0097	0.0150	-0.0035	-0.0045	0.0086	-0.1000	-0.0071
	(0.0105)	(0.0061)**	(0.0120)	(0.0111)	(0.0260)	(0.0898)	(0.0029)**
Fixed effects							
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry*Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.894	0.978	0.813	0.787	0.758	0.859	0.878
N	1,687	1,687	1,686	1,687	494	1,604	1,606

Panel A: H2b(2)	*Risk indicators (within-firm)*					*Firm value*	*Market's reaction*
	Leverage	Tangibility	Operating Profitability	Cash/Asset	Volatility of Cash/Asset	Tobin's q	Volatility of Market-Adjusted Return
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Treated * Post	0.0120	0.0010	0.0052	0.0018	0.0004	0.0057	-0.0058
	(0.0092)	(0.0053)	(0.0090)	(0.0092)	(0.0187)	(0.0827)	(0.0025)**
Fixed effects							
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry*Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.877	0.975	0.792	0.767	0.691	0.834	0.855
N	3,393	3,393	3,392	3,393	1,137	3,191	3,175

* p<0.1; ** p<0.05; *** p<0.01
Standard errors in parentheses

Appendix 2.1 Table 1
Variables definitions

Variable	Definition
Firm characteristics	
Leverage	The ratio of total debt to total assets.
Tangibility	The ratio of property, plant, and equipment (PPE) to total assets.
Operating profitability	The ratio the operating income before depreciation to total assets.
Tobin's q	Total assets minus the book value of equity plus the market value of equity, divided by the book value of total assets.
Total q	Aggregate market value scaled by total capital, the sum of the physical and intangible capital stocks.
Log(Sales)	The natural logarithm of sales.
Net equity issue	Change in book equity minus change balance sheet in retained earnings, divided by total assets.
Newly retained earnings	Change in balance sheet retained earnings, divided by total assets.
Asset growth	Change in total assets, divided by lagged total assets.
Net debt issue	Change in total assets minus change in book equity, divided by total assets.
Physical asset sales	Sale of PPE, divided by lagged total assets.
Asset turnover	Sales divided by total assets.
Cash / asset	The ratio of cash to total assets.
Volatility of cash / asset	The standard deviation of Cash/asset computed over 8 preceding quarters.
Volatility of market-adjusted return	Thestandard deviation (computed over 36 preceding months) of (the stock return less same period return on the CRSP value-weighted portfolio of NYSE/Amex/Nasdaq stocks).
Board characteristics	
Board size	The number directors on the board.
Female directors	The number of female directors on the board.
Fraction of female directors	The fraction of female directors is female directors, divided by board size.
Independent directors	The number of independent directors on the board.
Fraction of independent directors	Independent directors, divided by board size.
Female independent directors	The number of female independent directors
Fraction of female independent directors	The number of female independent directors, divided by independent directors.
Non-CEO leader	A leader on the board who is not the CEO (e.g., lead director).
Fraction of female non-CEO leaders	The number of female non-CEO leaders, divided by non-CEO leaders.
Female chairman	A dummy variable that takes the value if the chairman is female, and zero otherwise.
Female CEO	A dummy variable that takes the value if the CEO is female, and zero otherwise.
Female chairman / CEO	A dummy variable that takes the value if the chairman/CEO is female, and zero otherwise.

Appendix 2.2

Additional robustness tests

Table 3 – enlarged sample size[48]

Hypothesis test 1 (H1)

This table reports the stacked difference-in-differences regression results on the structured quasi-experimental treated and control samples corresponding to H1.

$$RI_{i,j,c,t} = \beta * Treated\_Post_{i,j,c,t} + p_t + m_c + \alpha_i + \vartheta_{j,t} + u_{i,j,c,t}$$

β captures the stacked difference-in-differences effect on each of the risk, firm value, or market reaction indicators ($RI_{i,j,c,t}$, named just above the column numbers, computed over the 3-year pre-event and 3-year post-event estimation windows for a firm i, in industry j, treated in year t, and belongs to cohort c). $Treated * Post$ is the interaction of the treated and post dummy variables and p_t, m_c, α_i, and $\vartheta_{j,t}$ are the year, cohort, firm, and industry-year fixed effects, respectively, that help minimize the heterogeneities across those dimensions. Leverage is the ratio of total debt to total assets. Tangibility is the ratio of PPE to total assets. Operating profitability is the ratio of the operating income before depreciation to total assets. Cash/asset is the ratio of cash to total assets. Volatility of the Cash/asset is the standard deviation of Cash/asset computed over 8 preceding quarters. Tobin's q is total assets minus the book value of equity plus the market value of equity, divided by the book value of total assets. Volatility of market-adjusted return is the standard deviation (computed over 36 preceding months) of the stock return less same period return on the CRSP value-weighted portfolio of NYSE/Amex/Nasdaq stocks.

	Risk indicators (within-firm)					*Firm value*	*Market's reaction*
	Leverage	Tangibility	Operating Profitability	Cash/Asset	Volatility of Cash/Asset	Tobin's q	Volatility of Market-Adjusted Return
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Treated * Post	-0.0167	-0.0039	-0.0017	0.0144	-0.0137	0.0751	0.0012
	(0.0107)	(0.0050)	(0.0067)	(0.0099)	(0.0174)	(0.0911)	(0.0024)
Fixed effects							
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry*Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.896	0.989	0.879	0.866	0.711	0.895	0.892
N	1,042	1,042	1,030	1,042	557	914	1,024

* p<0.1; ** p<0.05; *** p<0.01
Standard errors in parentheses

[48] Enlarged sample size simply means that we regenerate the same Tables in the main chapter by including observations from all firms in the utilities and financial industry that were previously excluded in the analyses, and we also drop the requirement for a treated (control) firm to have at least one firm from the same industry

Table 4 – enlarged sample size[49]
Hypothesis test 2a (H2a)
This table reports the stacked difference-in-differences regression results on the structured quasi-experimental treated and control samples corresponding to H2a.

$$RI_{i,j,c,t} = \beta * Treated\_Post_{i,j,c,t} + p_t + m_c + \alpha_i + \vartheta_{j,t} + u_{i,j,c,t}$$

β captures the stacked difference-in-differences effect on each of the risk, firm value, or market reaction indicators ($RI_{i,j,c,t}$, named just above the column numbers, computed over the 3-year pre-event and 3-year post-event estimation windows for a firm i, in industry j, treated in year t, and belongs to cohort c). $Treated * Post$ is the interaction of the treated and post dummy variables and p_t, m_c, α_i, and $\vartheta_{j,t}$ are the year, cohort, firm, and industry-year fixed effects, respectively, that help minimize the heterogeneities across those dimensions. Leverage is the ratio of total debt to total assets. Tangibility is the ratio of PPE to total assets. Operating profitability is the ratio of the operating income before depreciation to total assets. Cash/asset is the ratio of cash to total assets. Volatility of the Cash/asset is standard deviation of Cash/asset computed over 8 preceding quarters. Tobin's q is total assets minus the book value of equity plus the market value of equity, divided by the book value of total assets. Volatility of market-adjusted return is standard deviation (computed over 36 preceding months) of the stock return less same period return on the CRSP value-weighted portfolio of NYSE/AMEX/Nasdaq stocks.

	Risk indicators (within-firm)					*Firm value*	*Market's reaction*
	Leverage	Tangibility	Operating Profitability	Cash/Asset	Volatility of Cash/Asset	Tobin's q	Volatility of Market-Adjusted Return
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Treated * Post	-0.0432 (0.0099)***	-0.0018 (0.0063)	0.0350 (0.0066)***	0.0118 (0.0069)*	-0.0422 (0.0151)***	0.2246 (0.0771)***	-0.0009 (0.0022)
Fixed effects							
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry*Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.934	0.986	0.865	0.916	0.719	0.908	0.898
N	791	791	760	791	454	671	766

* p<0.1; ** p<0.05; *** p<0.01
Standard errors in parentheses

[49] Enlarged sample size simply means that we regenerate the same Tables in the main chapter by including observations from all firms in the utilities and financial industry that were previously excluded in the analyses, and we also drop the requirement for a treated (control) firm to have at least one firm from the same industry

Table 5 – enlarged sample size[50]
Are the H(2a) test results driven by equity market timing?
This table reports the stacked difference-in-differences regression results on the structured quasi-experimental treated and control samples corresponding to H2a. We add additional variables to evaluate whether the results in Table 2.4 are driven by equity market timing. The results are consistent with Rajan and Zingales (1995). However, they are not consistent with Baker and Wurgler (2002), suggesting that the effects of increasing power/influence of the female directors (via greater numerical strength) on leverage and its determinants are not coincident with equity market timing.

$$RI_{i,j,c,t} = \beta * Treated\_Post_{i,j,c,t} + p_t + m_c + \alpha_i + \vartheta_{j,t} + u_{i,j,c,t}$$

β captures the stacked difference-in-differences effect on each of the firm-level risk or equity market timing indicators ($RI_{i,j,c,t}$, named just above the column numbers, computed over the 3-year pre-event and 3-year post-event estimation windows for a firm i, in industry j, treated in year t, and belongs to cohort c). $Treated * Post$ is the interaction of the treated and post dummy variables and p_t, m_c, α_i, and $\vartheta_{j,t}$ are the year, cohort, firm, and industry-year fixed effects, respectively, that help minimize the heterogeneities across those dimensions. Leverage is the ratio of total debt to total assets. Tangibility is the ratio of PPE to total assets. Operating profitability is the ratio of the operating income before depreciation to total assets. Tobin's q is total assets minus the book value of equity plus the market value of equity, divided by the book value of total assets. Net debt issue is change in total assets minus change in book equity, divided by total assets. Net equity issue is change in book equity minus change in balance sheet retained earnings, divided by total assets. Newly retained earnings is change in balance sheet retained earnings, divided by total assets. Asset growth is change in total assets, divided by lagged total assets.

	H2a: Consistent with Zingales (1995)				*H2a: Not consistent with Baker and Wurgler (2002)*			
	Leverage	Tangibility	Operating Profitability	Tobin's q	Net Debt Issue	Net Equity Issue	Newly Retained Earnings	Asset Growth
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*
Treated * Post	-0.0432	-0.0018	0.0350	0.2246	-0.0250	0.0053	0.0265	-0.0071
	(0.0099)***	(0.0063)	(0.0066)***	(0.0771)***	(0.0216)	-0.0158	(0.0166)	-0.0962
Fixed effects								
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry*Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.934	0.986	0.865	0.908	0.524	0.573	0.508	0.645
N	791	791	760	671	543	543	656	656

* p<0.1; ** p<0.05; *** p<0.01
Standard errors in parentheses

[50] Enlarged sample size simply means that we regenerate the same Tables in the main chapter by including observations from all firms in the utilities and financial industry that were previously excluded in the analyses, and we also drop the requirement for a treated (control) firm to have at least one firm from the same industry

Table 6 – enlarged sample size[51]
Does activated managerial action drive the H(2a) test results?

This table reports the stacked difference-in-differences regression results on the structured quasi-experimental treated and control samples corresponding to H2a. We add additional variables to evaluate whether activated managerial actions drive the results in Table 2.4.

$$RI_{i,j,c,t} = \beta * Treated\_Post_{i,j,c,t} + p_t + m_c + \alpha_i + \vartheta_{j,t} + u_{i,j,c,t}$$

β captures the stacked difference-in-differences effect on each of the firm-level risk or managerial action indicators ($RI_{i,j,c,t}$, named just above the column numbers, computed over the 3-year pre-event and 3-year post-event estimation windows for a firm i, in industry j, treated in year t, and belongs to cohort c). $Treated * Post$ is the interaction of the treated and post dummy variables and p_t, m_c, α_i, and $\vartheta_{j,t}$ are the year, cohort, firm, and industry-year fixed effects, respectively, that help minimize the heterogeneities across those dimensions. Leverage is the ratio of total debt to total assets. Tangibility is the ratio of PPE to total assets. Operating profitability is the ratio of the operating income before depreciation to total assets. Tobin's q is total assets minus the book value of equity plus the market value of equity, divided by the book value of total assets. Physical asset sales is sale of PPE, divided by lagged total assets. Asset turnover is sales divided by total assets. Cash/asset is the ratio of cash to total assets. Newly retained earnings is change in balance sheet retained earnings, divided by total assets. Asset growth is change in total assets, divided by lagged total assets.

	H2a: Consistent with Zingales (1995)				*H2a: consistent with activated managerial action*			
	Leverage	Tangibility	Operating Profitability	Tobin's q	Physical Asset Sales	Asset Turnover	Cash/Asset	Newly Retained Earnings
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*
Treated * Post	-0.0432	-0.0018	0.0350	0.2246	0.0039	0.0678	0.0118	0.0265
	(0.0099)***	(0.0063)	(0.0066)***	(0.0771)***	(0.0017)**	(0.0395)*	(0.0069)*	(0.0166)
<u>Fixed effects</u>								
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry*Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.934	0.986	0.865	0.908	0.938	0.975	0.916	0.508
N	791	791	760	671	467	656	791	656

* p<0.1; ** p<0.05; *** p<0.01
Standard errors in parentheses

[51] Enlarged sample size simply means that we regenerate the same Tables in the main chapter by including observations from all firms in the utilities and financial industry that were previously excluded in the analyses, and we also drop the requirement for a treated (control) firm to have at least one firm from the same industry

Table 7 – enlarged sample size[52]
Hypothesis test 2b (H2b)

This table reports the stacked difference-in-differences regression results on the structured quasi-experimental treated and control samples corresponding to H2b. Panels A and B show the results for hypotheses 2b(1) and 2b(2), respectively. The main difference between hypotheses 2b(1) and 2b(2) is as follows. Hypothesis 2b(1) uses a control subsample involving male-to-male CEO transitions in firms with homogenous male boards. However, hypothesis 2b(2) uses a control subsample involving male-to-male CEO transitions in firms with gender-skewed majority male boards with one or more female directors.

$$RI_{i,j,c,t} = \beta * Treated\_Post_{i,j,c,t} + p_t + m_c + \alpha_i + \vartheta_{j,t} + u_{i,j,c,t}$$

β captures the stacked difference-in-differences effect on each of the risk, firm value, or market reaction indicators ($RI_{i,j,c,t}$, named just above the column numbers, computed over the 3-year pre-event and 3-year post-event estimation windows for a firm i, in industry j, treated in year t, and belongs to cohort c). $Treated * Post$ is the interaction of the treated and post dummy variables and p_t, m_c, α_i, and $\vartheta_{j,t}$ are the year, cohort, firm, and industry-year fixed effects, respectively, that help minimize the heterogeneities across those dimensions. Leverage is the ratio of total debt to total assets. Tangibility is the ratio of PPE to total assets. Operating profitability is the ratio of the operating income before depreciation to total assets. Cash/asset is the ratio of cash to total assets. Volatility of the Cash/asset is standard deviation of Cash/asset computed over 8 preceding quarters. Tobin's q is total assets minus the book value of equity plus the market value of equity, divided by the book value of total assets. Volatility of market-adjusted return is standard deviation (computed over 36 preceding months) of the stock return less same period return on the CRSP value-weighted portfolio of NYSE/AMEX/Nasdaq stocks.

Panel A: H2b(1)	*Risk indicators (within-firm)*					*Firm value*	*Market's reaction*
	Leverage	Tangibility	Operating Profitability	Cash/Asset	Volatility of Cash/Asset	Tobin's q	Volatility of Market-Adjusted Return
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Treated * Post	0.0089 (0.0083)	0.0109 (0.0046)**	-0.0061 (0.0090)	-0.0054 (0.0086)	0.0149 (0.0161)	-0.0841 (0.0781)	-0.0053 (0.0023)**
Fixed effects							
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry*Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.905	0.984	0.839	0.812	0.786	0.866	0.886
N	2,455	2,455	2,437	2,455	798	2,115	2,358

Panel A: H2b(2)	*Risk indicators (within-firm)*					*Firm value*	*Market's reaction*
	Leverage	Tangibility	Operating Profitability	Cash/Asset	Volatility of Cash/Asset	Tobin's q	Volatility of Market-Adjusted Return
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Treated * Post	0.0113 (0.0072)	0.0005 (0.0041)	0.0020 (0.0067)	-0.0019 (0.0072)	-0.0020 (0.0131)	0.0023 (0.0729)	-0.0058 (0.0020)***
Fixed effects							
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry*Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.894	0.980	0.824	0.792	0.713	0.846	0.866
N	5,141	5,141	5,103	5,141	1,752	4,423	4,839

* p<0.1; ** p<0.05; *** p<0.01
Standard errors in parentheses

[52] Enlarged sample size simply means that we regenerate the same Tables in the main chapter by including observations from all firms in the utilities and financial industry that were previously excluded in the analyses, and we also drop the requirement for a treated (control) firm to have at least one firm from the same industry

Appendix 2.IA. (Designated Internet Appendix).
Multi-Cohort Stacked DiD, Proof, and Simulation

In this appendix, we schematize the similarities/difference between single and multi-cohort stacked DiD, present the (i) proof of Proposition 1 and (ii) simulation results providing key insights into our multi-cohort stacked DiD analyses (on quasi-temporal counterfactual samples) over fixed, short, symmetric event windows.

2.IA.1. Stacked DiD: Single Cohort vs. Multi-Cohort

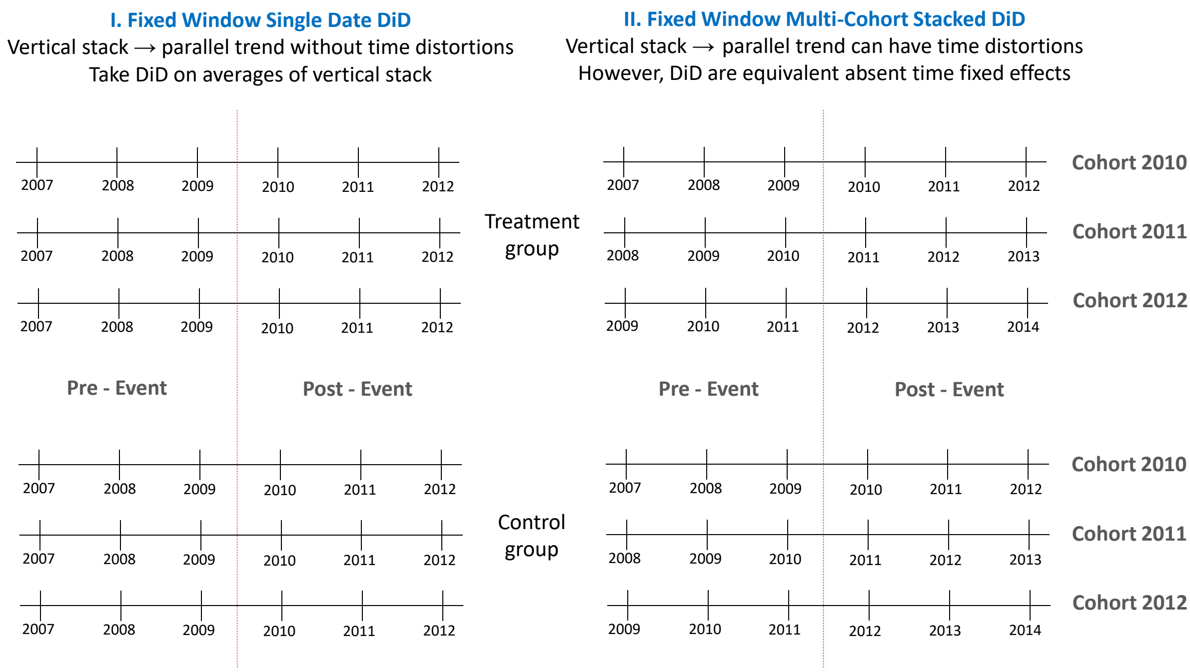


Figure 2.IA.1
This Figure shows the similarities and differences between a single-cohort (or single date) stacked DID analysis involving multiple units and the multi-cohort stacked DiD analysis. In both cases, treated and control samples from different firms are pooled according to their treatment conditions (i.e., treated group or control group) and aligned using their event date. In the single-date or single-cohort DiD analysis involving multiple firms, the event date is the same for all firms. The observations positioned at the same relative position from the event date have the same year of observation (y.o.b). Hence, the vertical stack of the treatment or control group will take the average of data points that have the same y.o.b. The latter will yield vertical stacks of treated and control groups that exhibit parallel trends without time distortions. However, in the multi-cohort stacked DiD analysis, the pooled and aligned firms belong to different cohorts with different event dates. In this case, the observations at the same relative position from the event date do not have the same y.o.b. The vertical stack of the treatment or control group will take the average of data points with different y.o.b., which will yield vertical stacks of treated and control groups that exhibit parallel trends with time distortions if time fixed effects are present. Nevertheless, in the absence of time fixed effects, both DiD analyses are equivalent.

2.IA.2. Proof of Proposition 1

Refer to the analytical model in section 2.3.4.

Suppose that the data are sorted by industry and by cohort. Suppose also that the treated and control firms per cohort are pooled and vertically stacked. To account for cases in which the quasi-temporal treated and control firms may have different firm fixed effects, let m_i^1 and m_i^0 be the average firm fixed effects for the pooled and stacked treated and control firms, respectively, per cohort. Let $\overline{y_{ijct}^{w,1}}$ and $\overline{y_{ijct}^{w,0}}$ be the average levels of the proxy of interest for the pooled treated and control samples, respectively, per cohort. Let also $\overline{\varepsilon_{ijct}^{w,1}}$ and $\overline{\varepsilon_{ijct}^{w,0}}$ be the average levels of the measurement errors for the pooled treated and control samples, respectively, per cohort. Then, without loss of generality, we have:

$$\overline{y_{ijct}^{w,1}} = y^{w,1} + m_i^1 + m_j + m_c + m_t + m_{ij}^1 + m_{ic}^1 + m_{it}^1 + m_{jc} + m_{jt} + m_{ct} + m_{ijc}^1 + m_{ijt}^1 + m_{ict}^1 + m_{jct} + m_{ijct}^1 + \overline{\varepsilon_{ijct}^{w,1}} \quad \textbf{(IA2.1)}$$

$$\overline{y_{ijct}^{w,0}} = y^{w,0} + m_i^0 + m_j + m_c + m_t + m_{ij}^0 + m_{ic}^0 + m_{it}^0 + m_{jc} + m_{jt} + m_{ct} + m_{ijc}^0 + m_{ijt}^0 + m_{ict}^0 + m_{jct} + m_{ijct}^0 + \overline{\varepsilon_{ijct}^{w,0}} \quad \textbf{(IA2.2)}$$

Let $\overline{RI}_{jc}^{DiD}$ be the DiD on vertical stacks computed per cohort at the industry level. By our symmetric construction of the event window, the sizes of the half-windows for the quasi-temporal counterfactual sample are equal. That is, $|W^+| = |W^-| = \frac{|W|}{2}$. Hence:

$$\overline{RI}_{jc}^{DiD} = \left[{}^{1}\!\Big/\!\frac{|W|}{2} \left(\sum_{w \epsilon W^+} \overline{y_{ijct}^{w,1}} - \sum_{w \epsilon W^-} \overline{y_{ijct}^{w,1}} \right) - {}^{1}\!\Big/\!\frac{|W|}{2} \left(\sum_{w \epsilon W^+} \overline{y_{ijct}^{w,0}} - \sum_{w \epsilon W^-} \overline{y_{ijct}^{w,0}} \right) \right]_{jc}$$

$$\frac{|W|}{2} * \overline{RI}_{jc}^{DiD} = \left[\left(\sum_{w \epsilon W^+} \overline{y_{ijct}^{w,1}} - \sum_{w \epsilon W^+} \overline{y_{ijct}^{w,0}} \right) - \left(\sum_{w \epsilon W^-} \overline{y_{ijct}^{w,1}} - \sum_{w \epsilon W^-} \overline{y_{ijct}^{w,0}} \right) \right]_{jc}$$

Hence:

$$\frac{|W|}{2} * \overline{RI}_{jc}^{DiD} = \left[\sum_{w \epsilon W^+} \left(\overline{y_{ijct}^{w,1}} - \overline{y_{ijct}^{w,0}} \right) - \sum_{w \epsilon W^-} \left(\overline{y_{ijct}^{w,1}} - \overline{y_{ijct}^{w,0}} \right) \right]_{jc}$$

From equations (A2.1) and (A2.2):

$$\overline{y_{\iota\jmath ct}^{w,1}} - \overline{y_{\iota\jmath ct}^{w,0}} = (y^{w,1} - y^{w,0}) + (m_i^1 - m_i^0) + (m_{ij}^1 - m_{ij}^0) + (m_{ic}^1 - m_{ic}^0) + (m_{it}^1 - m_{it}^0) + (m_{ijc}^1 - m_{ijc}^0) + (m_{ijt}^1 - m_{ijt}^0) + (m_{ict}^1 - m_{ict}^0) + (m_{ijct}^1 - m_{ijct}^0) + (\overline{\varepsilon_{\iota\jmath ct}^{w,1}} - \overline{\varepsilon_{\iota\jmath ct}^{w,0}}) \quad \textbf{(IA2.3)}$$

Putting A2.3 in $\frac{|W|}{2} * \overline{RI}_{jc}^{DiD}$, we have:

$$\frac{|W|}{2} * \overline{RI}_{jc}^{DiD} = \Bigg[\sum_{w\epsilon W^+} \Big((y^{w,1} - y^{w,0}) + (m_i^1 - m_i^0) + (m_{ij}^1 - m_{ij}^0) + (m_{ic}^1 - m_{ic}^0) + (m_{it}^1 - m_{it}^0) + (m_{ijc}^1 - m_{ijc}^0) + (m_{ijt}^1 - m_{ijt}^0) + (m_{ict}^1 - m_{ict}^0) + (m_{ijct}^1 - m_{ijct}^0) + (\overline{\varepsilon_{\iota\jmath ct}^{w,1}} - \overline{\varepsilon_{\iota\jmath ct}^{w,0}})\Big) - \sum_{w\epsilon W^-} \Big((y^{w,1} - y^{w,0}) + (m_i^1 - m_i^0) + (m_{ij}^1 - m_{ij}^0) + (m_{ic}^1 - m_{ic}^0) + (m_{it}^1 - m_{it}^0) + (m_{ijc}^1 - m_{ijc}^0) + (m_{ijt}^1 - m_{ijt}^0) + (m_{ict}^1 - m_{ict}^0) + (m_{ijct}^1 - m_{ijct}^0) + (\overline{\varepsilon_{\iota\jmath ct}^{w,1}} - \overline{\varepsilon_{\iota\jmath ct}^{w,0}})\Big)\Bigg]_{jc}$$

Again, by the ex-ante sorting, we have:

$$\frac{|W|}{2} * \overline{RI}_{jc}^{DiD} = \Bigg[\sum_{w\epsilon W^+} \Big((y^{w,1} - y^{w,0}) + (m_{it}^1 - m_{it}^0) + (m_{ijt}^1 - m_{ijt}^0) + (m_{ict}^1 - m_{ict}^0) + (m_{ijct}^1 - m_{ijct}^0) + (\overline{\varepsilon_{\iota\jmath ct}^{w,1}} - \overline{\varepsilon_{\iota\jmath ct}^{w,0}})\Big) - \sum_{w\epsilon W^-} \Big((y^{w,1} - y^{w,0}) + (m_{it}^1 - m_{it}^0) + (m_{ijt}^1 - m_{ijt}^0) + (m_{ict}^1 - m_{ict}^0) + (m_{ijct}^1 - m_{ijct}^0) + (\overline{\varepsilon_{\iota\jmath ct}^{w,1}} - \overline{\varepsilon_{\iota\jmath ct}^{w,0}})\Big)\Bigg]_{jc}$$

Hence:

$$\frac{|W|}{2} * \overline{RI}_{jc}^{DiD} = \Bigg[\sum_{w\epsilon W^+}(y^{w,1} - y^{w,0}) - \sum_{w\epsilon W^-}(y^{w,1} - y^{w,0})$$
$$+ \sum_{w\epsilon W^+}(m_{it}^1 - m_{it}^0) - \sum_{w\epsilon W^-}(m_{it}^1 - m_{it}^0)$$
$$+ \sum_{w\epsilon W^+}\left(m_{ijt}^1 - m_{ijt}^0\right) - \sum_{w\epsilon W^-}\left(m_{ijt}^1 - m_{ijt}^0\right)$$
$$+ \sum_{w\epsilon W^+}(m_{ict}^1 - m_{ict}^0) - \sum_{w\epsilon W^-}(m_{ict}^1 - m_{ict}^0)$$
$$+ \sum_{w\epsilon W^+}\left(m_{ijct}^1 - m_{ijct}^0\right) - \sum_{w\epsilon W^-}\left(m_{ijct}^1 - m_{ijct}^0\right)$$
$$+ \sum_{w\epsilon W^+}\left(\overline{\varepsilon_{\imath\jmath ct}^{w,1}} - \overline{\varepsilon_{\imath\jmath ct}^{w,0}}\right) - \sum_{w\epsilon W^-}\left(\overline{\varepsilon_{\imath\jmath ct}^{w,1}} - \overline{\varepsilon_{\imath\jmath ct}^{w,0}}\right)\Bigg]_{jc}$$

Without loss of generality, by the ex-ante sorting, we also have:

$$\sum_{w\epsilon W^+}(m_{it}^1 - m_{it}^0) - \sum_{w\epsilon W^-}(m_{it}^1 - m_{it}^0) + \sum_{w\epsilon W^+}\left(m_{ijt}^1 - m_{ijt}^0\right) - \sum_{w\epsilon W^-}\left(m_{ijt}^1 - m_{ijt}^0\right)$$
$$+ \sum_{w\epsilon W^+}(m_{ict}^1 - m_{ict}^0) - \sum_{w\epsilon W^-}(m_{ict}^1 - m_{ict}^0)$$
$$+ \sum_{w\epsilon W^+}\left(m_{ijct}^1 - m_{ijct}^0\right) - \sum_{w\epsilon W^-}\left(m_{ijct}^1 - m_{ijct}^0\right) = \underbrace{\rho_t(m_i^1 - m_i^0)}_{\to 0}$$

Where the value of ρ_t depends on the anti-symmetry of the distribution of the time fixed effects around the event dates for each cohort (again, at the industry level). Hence:

$$\frac{|W|}{2} * \overline{RI}_{jc}^{DiD} = \Bigg[\sum_{w\epsilon W^+}(y^{w,1} - y^{w,0}) - \sum_{w\epsilon W^-}(y^{w,1} - y^{w,0}) + \rho_t(m_i^1 - m_i^0)$$
$$+ \sum_{w\epsilon W^+}\left(\overline{\varepsilon_{\imath\jmath ct}^{w,1}} - \overline{\varepsilon_{\imath\jmath ct}^{w,0}}\right) - \sum_{w\epsilon W^-}\left(\overline{\varepsilon_{\imath\jmath ct}^{w,1}} - \overline{\varepsilon_{\imath\jmath ct}^{w,0}}\right)\Bigg]_{jc}$$

$$\overline{RI}_{jc}^{DiD} = \frac{2}{|W|}\left[\sum_{w\epsilon W^+}(y^{w,1} - y^{w,0}) - \sum_{w\epsilon W^-}(y^{w,1} - y^{w,0})\right]_{jc}$$
$$+ \frac{2}{|W|}\left[\rho_t(m_i^1 - m_i^0) + \left(\sum_{w\epsilon W^+}\left(\overline{\varepsilon_{\imath\jmath ct}^{w,1}} - \overline{\varepsilon_{\imath\jmath ct}^{w,0}}\right) - \sum_{w\epsilon W^-}\left(\overline{\varepsilon_{\imath\jmath ct}^{w,1}} - \overline{\varepsilon_{\imath\jmath ct}^{w,0}}\right)\right)\right]_{jc}$$

Stacking the DiDs computed at the cohort and industry level sequentially, we have:

$$\overline{RI}^{DiD} = \frac{1}{N_j * N_c} \sum_j \sum_c \overline{RI}_{jc}^{DiD}$$

Where $\overline{RI}^{DiD}$ is the stacked DiD, N_j (N_c) is the number of industries (cohorts). Hence:

$$\overline{RI}^{DiD} = \underbrace{\frac{2}{|W|N_j * N_c} \sum_j \sum_c \left[\sum_{w \epsilon W^+} (y^{w,1} - y^{w,0}) - \sum_{w \epsilon W^-} (y^{w,1} - y^{w,0}) \right]_{jc}}_{Vertical\ stack\ of\ true\ DiD}$$
$$+ \frac{2}{|W|N_j * N_c} \sum_j \sum_c \Bigg[\rho_t (m_i^1 - m_i^0)$$
$$+ \left(\sum_{w \epsilon W^+} \left(\overline{\varepsilon_{\iota jct}^{w,1}} - \overline{\varepsilon_{\iota jct}^{w,0}} \right) - \sum_{w \epsilon W^-} \left(\overline{\varepsilon_{\iota jct}^{w,1}} - \overline{\varepsilon_{\iota jct}^{w,0}} \right) \right) \Bigg]_{jc}$$

Since the true data are by definition indifferent in the absence of fixed effects, then:

$$\overline{RI}^{DiD} = \underbrace{\frac{2}{|W|} \left(\sum_{w \epsilon W^+} (y^{w,1} - y^{w,0}) - \sum_{w \epsilon W^-} (y^{w,1} - y^{w,0}) \right)}_{True\ DiD}$$
$$+ \underbrace{\frac{2}{|W|N_j * N_c} \sum_j \sum_c [\rho_t (m_i^1 - m_i^0)]_{jc}}_{Vertical\ stack\ of\ anti-symmetry\ of\ fixed\ effects \to 0}$$
$$+ \underbrace{\frac{2}{|W|N_j * N_c} \sum_j \sum_c \left[\sum_{w \epsilon W^+} \left(\overline{\varepsilon_{\iota jct}^{w,1}} - \overline{\varepsilon_{\iota jct}^{w,0}} \right) - \sum_{w \epsilon W^-} \left(\overline{\varepsilon_{\iota jct}^{w,1}} - \overline{\varepsilon_{\iota jct}^{w,0}} \right) \right]_{jc}}_{Vertical\ stack\ of\ the\ DiD\ on\ the\ measurement\ error \to 0}$$

By assumption, given the numerosity of the number of firms per cohort, the number of cohorts per industry, it follows that the vertical stack of the anti-symmetry of fixed effects around the event dates and the vertical stack of the DiD on the residual error term are both extremely close to zero. Hence:

$$\overline{RI}^{DiD} \cong \underbrace{\frac{2}{|W|}\left(\sum_{w\epsilon W^{+}}(y^{w,1}-y^{w,0})-\sum_{w\epsilon W^{-}}(y^{w,1}-y^{w,0})\right)}_{True\ DiD}$$

If the vertical stack of the anti-symmetry of fixed effects around the event dates and the vertical stack of the DiD on the residual error term are both precisely zero, then:

$$\overline{RI}^{DiD} = \frac{2}{|W|N_j * N_c}\sum_{j}\sum_{c}\left[\left(\sum_{w\epsilon W^{+}}\overline{y_{\iota\jmath ct}^{w,1}}-\sum_{w\epsilon W^{-}}\overline{y_{\iota\jmath ct}^{w,1}}\right)-\left(\sum_{w\epsilon W^{+}}\overline{y_{\iota\jmath ct}^{w,0}}-\sum_{w\epsilon W^{-}}\overline{y_{\iota\jmath ct}^{w,0}}\right)\right]_{jc}$$

$$= \underbrace{\frac{2}{|W|}\left(\sum_{w\epsilon W^{+}}(y^{w,1}-y^{w,0})-\sum_{w\epsilon W^{-}}(y^{w,1}-y^{w,0})\right)}_{True\ DiD}$$

■

2.IA.3. Model and Data

Suppose there exist true data for an arbitrary pair of quasi-temporal counterfactual events[53] defined as follows. For all discrete observation dates, ω, in the event window, W, the true level of a risk indicator is $y^{w,T}$, where $T = \{0, 1\}$ is the set of treatment conditions such that $T = 1$ corresponds to the treated group and $T = 0$ to the counterfactual or control group. Let $W = \{t-3,\ t-2,\ t-1,\ t,\ t+1,\ t+2\}$, such that the post-event half-window, $W^{+} = \{t,\ t+1,\ t+2\}$ and the pre-event half-window, $W^{-} = \{t-3,\ t-2,\ t-1\}$. Set:

$$\underbrace{y^{w,T}}_{T=\{0,1\},\ w\epsilon W} = \begin{bmatrix} y^{t-3,1} & y^{t-3,0} \\ y^{t-2,1} & y^{t-2,0} \\ y^{t-1,1} & y^{t-1,0} \\ y^{t,1} & y^{t,0} \\ y^{t+1,1} & y^{t+1,0} \\ y^{t+2,1} & y^{t+2,0} \end{bmatrix} = \begin{bmatrix} 10 & 10 \\ 10 & 10 \\ 10 & 10 \\ 9 & 14 \\ 9 & 14 \\ 9 & 14 \end{bmatrix}$$

Such that the levels of the risk indicator or proxy variable of interest for the treatment and control groups (i.e., $y^{w,1}$ and $y^{w,0}$, respectively), are indistinguishable in the pre-event window, W^{-}, but diverge in the post-event window, W^{+} (with $y^{w,1}$ falling from 10 to 9, and $y^{w,0}$ rising from 10 to 14; all in the relevant units of measurement). We plot the true image of $\underbrace{y^{w,T}}_{T=\{0,1\},\ w\epsilon W}$ and the "DiD on the true image" (or true DiD) in Figures 2.IA.3.1 (a) and (b):

[53] Recall that our quasi-temporal counterfactual events represent a pair of defined counterfactual conditions in the post-event period. Take the first of our four hypotheses in the main text (i.e., H1) as a case in point. All boards are homogenous-male boards, each with a male CEO in the pre-event period, $W^{-} = \{t-3,\ t-2,\ t-1\}$. However, in the post-event period, $W^{+} = \{t,\ t+1,\ t+2\}$, the defined counterfactual conditions are as follows. For one group of the boards, each board adds a female director in a non-leading role, all else equal, and for the other group of boards, each board adds a male director in a non-leading role, with all else unchanged. The first group is equivalent to the treatment (or treated) group, and the second is the counterfactual (or control) group. Ideally (e.g., as in H1, H2a, and H2b(2)) but not compulsorily (e.g., as in H2b(1)), the quasi-temporal events would be indistinguishable in the pre-event period.

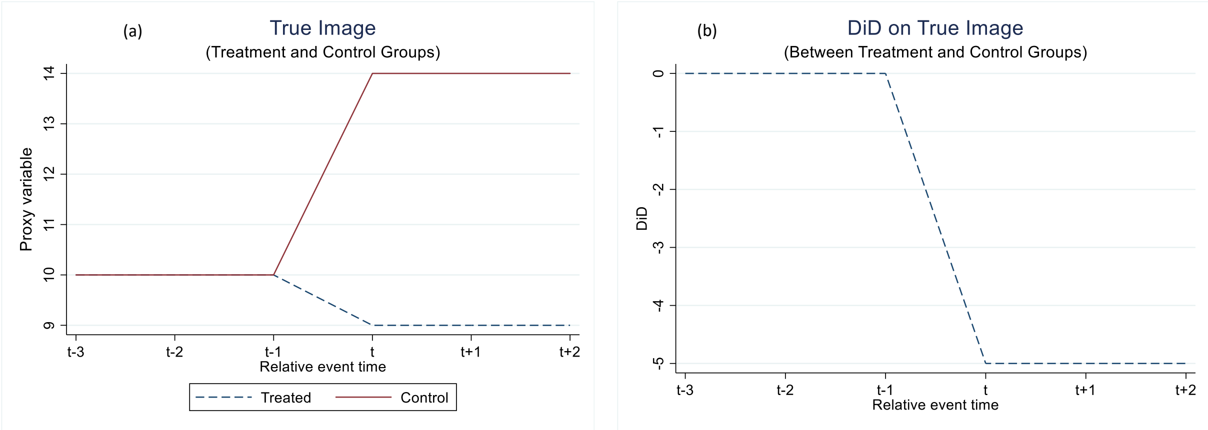


Figure 2.IA.3.1. True image / DiD of proxy variable for the set of quasi-temporal samples

Suppose further that when a sample of observations drawn from a firm satisfies either treatment condition (i.e., treated or control), the real differences between the measured levels of such observations and those of the true data are due principally to the full suite of fixed effects associated with the firm, the cohort and industry to which the firm belongs, and the time (e.g., year) of the observations. Then, symbolically:

$$y_{ijct}^{w,T} = y^{w,T} + m_i + m_j + m_c + m_t + m_{ij} + m_{ic} + m_{it} + m_{jc} + m_{jt} + m_{ct} + m_{ijc} + m_{ijt} + m_{ict} + m_{jct} + m_{jct} + m_{ijct} + \varepsilon_{ijct}^{w,T} \quad \textbf{(IA.3.1)}$$

Where $y_{ijct}^{w,T}$ is the observed level of the risk indicator measured at firm i in industry j and cohort c at time t (corresponding to the relative observation date w) with treatment condition T. $y^{w,T}$ is the true level of the risk indicator for a quasi-counterfactual sample with treatment condition T at relative observation date w. m_p is the p fixed-effect and p is the full suite of individual or interactive fixed effects. Without sorting, it follows that:

$$y_{ijct}^{w,1} = y^{w,1} + m_i^1 + m_j^1 + m_c^1 + m_t^1 + m_{ij}^1 + m_{ic}^1 + m_{it}^1 + m_{jc}^1 + m_{jt}^1 + m_{ct}^1 + m_{ijc}^1 + m_{ijt}^1 + m_{ict}^1 + m_{jct} + m_{ijct}^1 + \varepsilon_{ijct}^{w,1} \quad \textbf{(IA.3.2)}$$

$$y_{ijct}^{w,0} = y^{w,0} + m_i^0 + m_j^0 + m_c^0 + m_t^0 + m_{ij}^0 + m_{ic}^0 + m_{it}^0 + m_{jc}^0 + m_{jt}^0 + m_{ct}^0 + m_{ijc}^0 + m_{ijt}^0 + m_{ict}^0 + m_{jct} + m_{ijct}^0 + \varepsilon_{ijct}^{w,0} \quad \textbf{(IA.3.3)}$$

Let $c = 1, 2, \dots, C_i$ number of cohorts per firm; $t = 1, 2, \dots, T_i$ number of firm-year observations per firm; $i = 1, 2, \dots, I_j$ number of firms per industry; and $j = 1, 2, \dots, J_s$ number of industries in the simulation sample. We have $|W| = 6$. Let $T_i = 20, J_s = 10$, and, for simplicity, fix $C_i = 1$, such that a firm either contains only one treatment or one control sample. Let every observation year be associated with at least one treatment and one control sample, and let every industry have one treated and one control firm. Thus, the total number of treated and control cohorts per industry are $C_j^1 = T_i - |W| + 1 = 20 - 6 + 1 = 15 = C_j^0$ and the total number of firms per industry, $I_j = 2 * C_j^1 = 2 * C_j^0 = 30$. Hence, post-search, the total number of firm-year observations corresponding to the set of quasi-temporal counterfactual samples for the simulation exercise is $N = |W| * C_i * I_j * J_s = 6 * 1 * 30 * 10 = 1800$. Let each fixed effect, m_p, be distributed $\alpha_{m_p} * N(0, \sigma^2{}_{m_p})$ over the appropriate full length, where $\sigma^2{}_{m_p}$ is the variance of m_p; and α_{m_p} is a scaling factor that modulates the direction and severity of the heterogeneities due to each fixed effect, m_p.

2.IA.4. Simulating the Impact of the Fixed Effects

To simulate the impact of the various fixed effects on our multi-cohort stacked DiD (on structured quasi-temporal counterfactual samples) analyses over fixed, short, symmetric event windows, we look at several diagnostic objects: the vertical stacks, the normalized vertical stacks, and the normalized DiD. We define a vertical stack of the treated or control samples as gross averaging across units (i.e., firms), as follows:

$$RI_{\omega,\ Treated}^{Stacked} = 1\{W = \omega,\ \ Treated = 1\} * \frac{1}{N_1} \sum_{i=1}^{N_1} RI_i \qquad \textbf{(IA.4.4)}$$

$$RI_{\omega,\ Control}^{Stacked} = 1\{W = \omega,\ \ Treated = 0\} * \frac{1}{N_0} \sum_{i=1}^{N_0} RI_i \qquad \textbf{(IA.4.5)}$$

Where RI is a risk indicator or proxy variable of interest. N_1 and N_0 are the number of units (firms, in the context of this study) in the treated and control groups, respectively. $W = \{t-3,\ t-2,\ t-1,\ t,\ t+1,\ t+2\}$. $Treated = \{0, 1\}$ where one (1) corresponds to a treated sample and zero (0) corresponds to a control sample. We have $N_1 = N_0 = 900$.

To compute the normalized vertical stack of the treated/control samples, first, we calculate the half-window averages of the vertical stacks ($\overline{RI_{\omega \epsilon W^-,\ Treated}^{Stacked}}$; $\overline{RI_{\omega \epsilon W^+,\ Treated}^{Stacked}}$) for the treated group and ($\overline{RI_{\omega \epsilon W^-,\ Control}^{stacked}}$; $\overline{RI_{\omega \epsilon W^+,\ Control}^{Stacked}}$) for the control group. Finally, we adjust the stacks by the difference between $\overline{RI^{Stacked}_{\omega\epsilon W^-,\ Treated}}$ and $\overline{RI^{stacked}_{\omega\epsilon W^-,\ Control}}$. The goal is to reduce the average pre-event window difference to zero. The advantage is that the average post-event window difference between the treated and control groups easily corresponds to the estimated DiD, $\overline{RI}^{DiD}$, on the vertical stacks. Mathematically:

$$\overline{RI}^{Difference}_{Treated} = \overline{RI^{Stacked}_{\omega\epsilon W^+,\ Treated}} - \overline{RI^{Stacked}_{\omega\epsilon W^-,\ Treated}} \qquad \textbf{(IA.4.6)}$$

$$\overline{RI}^{Difference}_{Control} = \overline{RI^{Stacked}_{\omega\epsilon W^+,\ Control}} - \overline{RI^{stacked}_{\omega\epsilon W^-,\ Control}} \qquad \textbf{(IA.4.7)}$$

$$\overline{RI}^{DiD} = \overline{RI}^{Difference}_{Treated} - \overline{RI}^{Difference}_{Control} \qquad \textbf{(IA.4.8)}$$

Where RI is the risk indicator or proxy variable of interests. $\overline{RI}^{DiD}$ is the average DiD computed for the risk indicator. $W^- = \{t-3,\ t-2,\ t-1\}$ and $W^+ = \{t,\ t+1,\ t+2\}$.

2.IA.4.1. Time fixed effects (A)

Figure 2.IA.2 compares the vertical stacks of the true data and their DiD to the vertical stacks of the simulated measured data and their DiD, $\overline{RI}^{DiD}$, when the main difference between the measured data and the true data is the presence of time fixed effects. Mathematically:

$$y^{w,1}_{ijct} = y^{w,1} + m^1_t + \varepsilon^{w,1}_{ijct}$$

$$y^{w,0}_{ijct} = y^{w,0} + m^0_t + \varepsilon^{w,0}_{ijct}$$

Where m_t is distributed $\alpha_{m_t} * N(0, \sigma^2{}_{m_t})$ over T_i; $\sigma^2{}_{m_t}$ is the variance of m_t; and α_{m_t} is a scaling factor that modulates the direction and severity of the heterogeneities due to m_t.

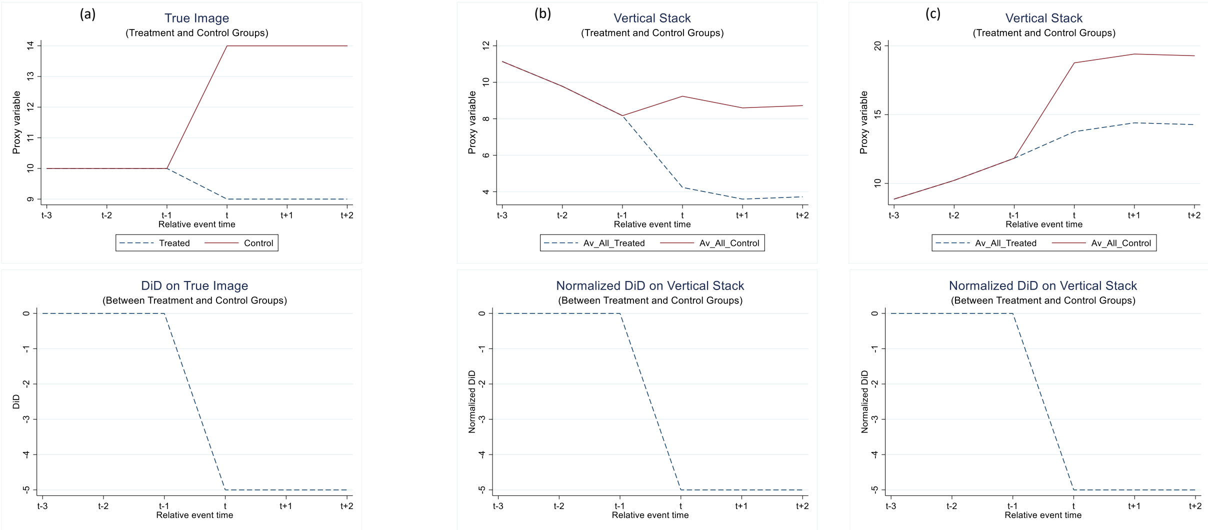


Figure 2.IA.4.1. Effect of time fixed effects on the image of the vertical stack

The top and bottom of column (a) show the vertical stacks of the true data and their DiD, respectively. The top of column (b) and the top column (c) show how time fixed effects can severely distort the image of vertical stack of the true data. The distribution of the time fixed effects are unchanged in both cases (i.e., images b and c). The only difference is that $\alpha_{m_t} = 8$ in column (b) and $\alpha_{m_t} = -8$ in column (c). Nevertheless, the bottom images in columns (b) and (c) show that, irrespective of the severity of time distortions, the normalized DiD on the vertical stack robustly recovers the DiD of the true image. Note that the recovery is exact since the true data for this simulation is devoid of noise.

2.IA.4.2. Time fixed effects (B)

Figure 2.IA.3 compares the vertical stacks of the true data and their DiD to the vertical stacks of the simulated measured data and their DiD, $\overline{RI}^{DiD}$, when the main difference between the measured data and the true data is the presence of time fixed effects. Mathematically:

$$y_{ijct}^{w,1} = y^{w,1} + m_t^1 + \varepsilon_{ijct}^{w,1}$$

$$y_{ijct}^{w,0} = y^{w,0} + m_t^0 + \varepsilon_{ijct}^{w,0}$$

Where m_t is distributed $\alpha_{m_t} * \widetilde{N}(k, \delta^2{}_{m_t})$ over T_i; $\delta^2{}_{m_t}$ is the variance of m_t; and α_{m_t} is a scaling factor that modulates the direction and severity of the heterogeneities due to m_t. $k \neq 0$ and $\widetilde{N}(k, \delta^2{}_{m_t})$ is only approximately normal. The goal of this additional exercise is to further illustrate the robustness of the recovery of the DiD.

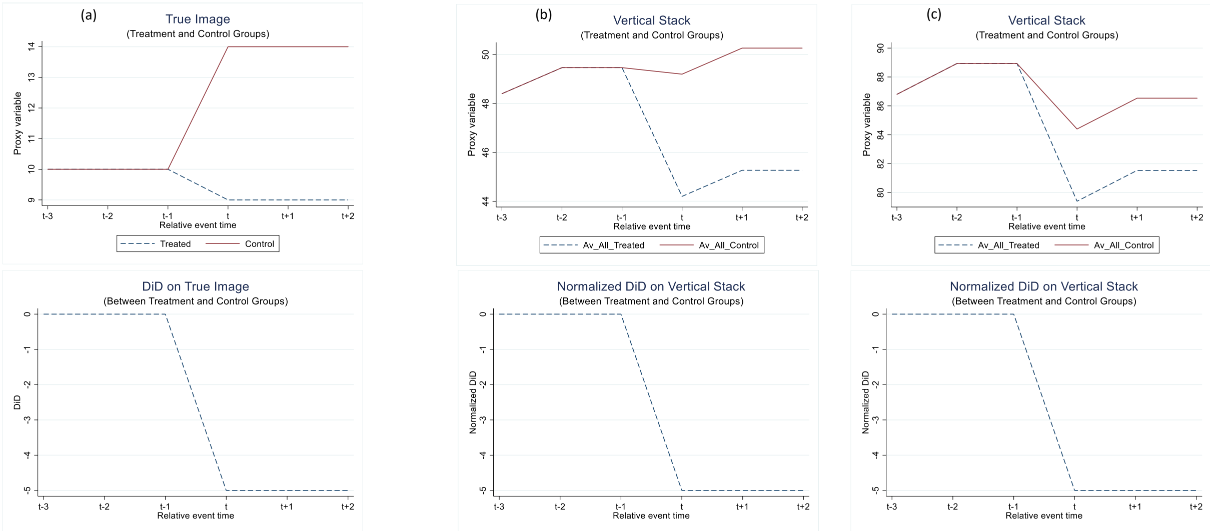


Figure 2.IA.4.2. Effect of time fixed effects on the image of the vertical stack

The top and bottom of column (a) show the vertical stacks of the true data and their DiD, respectively. The top of column (b) and the top column (c) show how time fixed effects with a different distributional can severely distort the image of vertical stack of the true data. The distribution of the time fixed effects are unchanged in both cases (i.e., images b and c). The only difference is that $\alpha_{m_t} = 2$ in column (b) and $\alpha_{m_t} = 4$ in column (c). Nevertheless, the bottom images in columns (b) and (c) show that, irrespective of the severity of time distortions, the normalized DiD on the vertical stack robustly recovers the DiD of the true image. Note that the recovery is exact since the true data for this simulation is devoid of noise.

2.IA.4.3. Firm fixed effects

Figure 2.IA.3 compares the vertical stacks of the true data and their DiD to the vertical stacks of the simulated measured data and their DiD, $\overline{RI}^{DiD}$, when the main difference between the measured data and the true data is the presence of firm fixed effects. Mathematically:

$$y_{ijct}^{w,1} = y^{w,1} + m_i^1 + \varepsilon_{ijct}^{w,1}$$

$$y_{ijct}^{w,0} = y^{w,0} + m_i^0 + \varepsilon_{ijct}^{w,0}$$

Where m_i is distributed $\alpha_i * N(0, \sigma^2{}_{m_i})$ over all firms; $\sigma^2{}_{m_i}$ is the variance of m_i; and α_{m_i} is a scaling factor that modulates the direction and severity of the heterogeneities due to m_i.

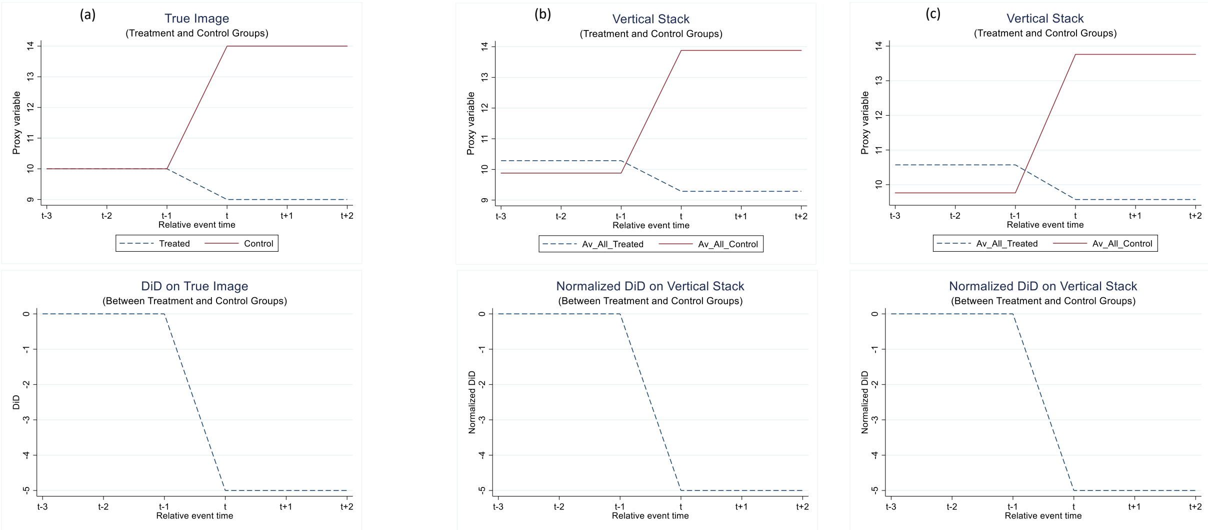


Figure 2.IA.4.3. Effect of firm fixed effects on the image of the vertical stack

The top and bottom of column (a) show the vertical stacks of the true data and their DiD, respectively. The top of column (b) and the top column (c) show that firm fixed effects, alone, only induce a cosmetic separation between the average levels for the treatment and the control groups (quite obvious in the pre-event window). $N(0, \sigma^2{}_{m_i})$ is unchanged in both cases (i.e., images b and c). The only difference is that $\alpha_{m_i} = 4$ in column (b) and $\alpha_{m_i} = 8$ in column (c). The latter implies that the larger the heterogeneities across firms, the higher the cosmetic separation. Nevertheless, the bottom images in columns (b) and (c) show that, irrespective of the size of the cosmetic separation, the normalized DiD on the vertical stack robustly recovers the DiD of the true image. Note that the recovery is exact since the true data for this simulation is devoid of noise.

2.IA.4.4. Combining time fixed effects (A) and the firm fixed effects above

Figure 2.IA.5 compares the vertical stacks of the true data and their DiD to the vertical stacks of the simulated measured data and their DiD, $\overline{RI}^{DiD}$, when the main difference between the measured data and the true data is the presence of the time fixed effects (A) and the firm fixed effects discussed above. Mathematically:

$$y_{ijct}^{w,1} = y^{w,1} + m_i^1 + m_t^1 + \varepsilon_{ijct}^{w,1}$$

$$y_{ijct}^{w,0} = y^{w,0} + m_i^0 + m_t^0 + \varepsilon_{ijct}^{w,0}$$

Where m_p is distributed $\alpha_p * N(0, \sigma^2{}_{m_p})$ as previously discussed and α_p is the same scaling factor as in the previous exercises, respectively.

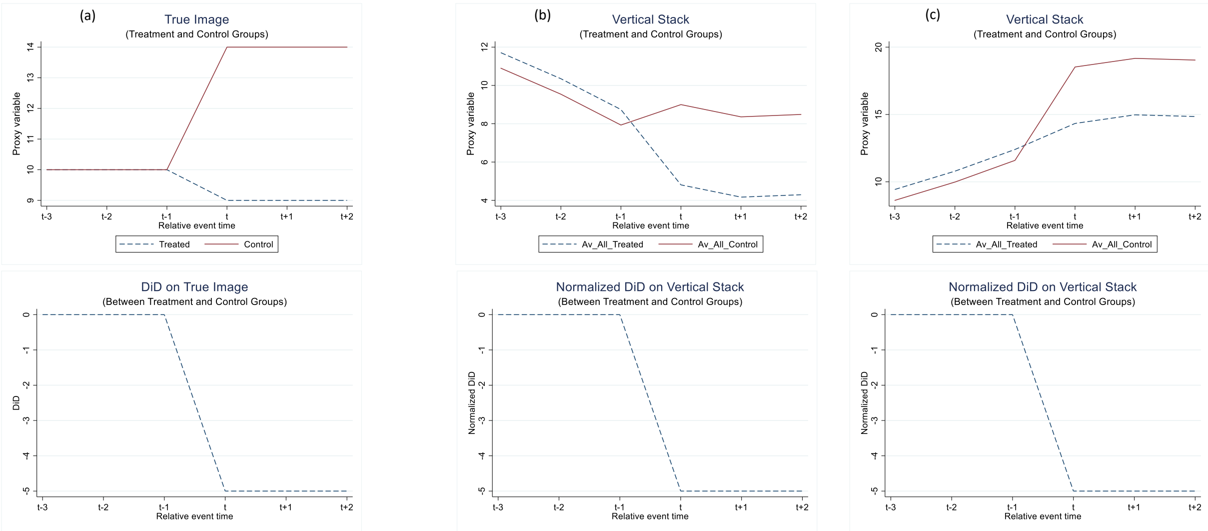


Figure 2.IA.4.4. Effect of time and firm fixed effects on the image of the vertical stack

The top and bottom of column (a) show the vertical stacks of the true data and their DiD, respectively. The top of column (b) and the top column (c) show the impact of the combination of the time and firm fixed effects as previously discussed. We observe both the distortion due to time fixed effects and separation (quite obvious in the pre-event window) due to firm fixed effects. Nevertheless, the bottom images in columns (b) and (c) show that, irrespective of these combined effects on the vertical stacks, the normalized DiD on the vertical stacks still robustly recovers the DiD of the true image. Note that the recovery is exact since the true data for this simulation is devoid of noise.

2.IA.4.5. Cohort fixed effects

Figure 2.IA.6 compares the vertical stacks of the true data and their DiD to the vertical stacks of the simulated measured data and their DiD, $\overline{RI}^{DiD}$, when the main difference between the measured data and the true data is the presence of cohort fixed effects. Mathematically:

$$y_{ijct}^{w,1} = y^{w,1} + m_c^1 + \varepsilon_{ijct}^{w,1}$$

$$y_{ijct}^{w,0} = y^{w,0} + m_c^0 + \varepsilon_{ijct}^{w,0}$$

Where m_c is distributed $\alpha_{m_c} * N(0, \sigma^2{}_{m_t})$ over C_j; $\sigma^2{}_{m_c}$ is the variance of m_c; and α_{m_c} is a scaling factor that modulates the direction and severity of the heterogeneities due to m_c.

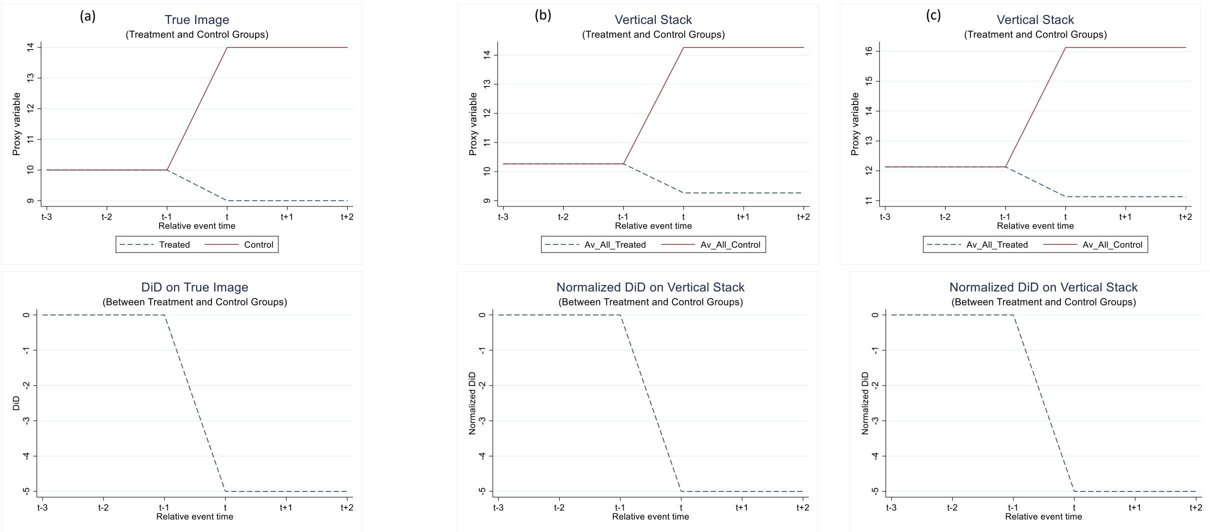


Figure 2.IA.4.5. Effect of cohort fixed effects on the image of the vertical stack

The top and bottom of column (a) show the vertical stacks of the true data and their DiD, respectively. The top of column (b) and the top column (c) show that cohort fixed effects do not distort the image of the vertical stack of the true data. Instead, the impact of the cohort fixed effect is to bulk-shift the average levels for the treatment and the control groups in the pre-event window. $N(0, \sigma^2{}_{m_c})$ is unchanged in both cases (i.e., images b and c). The only difference is that $\alpha_{m_c} = 2$ in column (b) and $\alpha_{m_c} = 8$ in column (c). The latter implies that the larger the heterogeneities across cohorts, the larger the bulk shift. Nevertheless, the bottom images in columns (b) and (c) show that, irrespective of the bulk shift, the normalized DiD on the vertical stack robustly recovers the DiD of the true image. Note that the recovery is exact since the true data for this simulation is devoid of noise.

2.IA.4.6. Other fixed effects and the robustness of the functional regression model

We also find that the industry fixed effects impact the vertical stacks by bulk-shifting the treated and control samples (similar to the impact of the cohort fixed effects). Next, we create data with different combinations of the fixed effects (up to the full suite of fixed effects) to evaluate whether our adaption of the Bertrand and Mullainathan (2003) functional regression model produces highly robust empirical estimates.

$$RI_{i,j,c,t} = \beta * Treated\_Post_{i,j,c,t} + p_t + m_c + \alpha_i + \vartheta_{j,t} + u_{i,j,c,t}$$

Where β captures the stacked difference-in-differences effect on each of the risk indicators ($RI_{i,j,c,t}$, described above, computed over the 3-year pre-event and 3-year post-event estimation windows for a firm i, in industry j, treated in year t, and belongs to cohort c). $Treated\_Post$ is the interaction of the treated and post dummy variables and p_t, m_c, α_i, and $\vartheta_{j,t}$ are the year, cohort, firm, and industry-year fixed effects, respectively, that help minimize the heterogeneities across those dimensions. We, below, show the results.

	True value	*True value + time FE*	*True value + firm FE*	*True value + time FE + firm FE*	*True value + time FE + firm FE + cohort FE*	*True value + time FE + firm FE + cohort FE + Industry FE*	*True value + time FE + firm FE + cohort FE + Industry FE + Industry-Year FE*
Treated * Post	-5	-5	-5	-5	-5	-5	-5
	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)***
Fixed effects							
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry_year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	1	1	1	1	1	1	1
N	1,800	1,800	1,800	1,800	1,800	1,800	1,800

* p <0.1; ** p <0.05; *** p <0.01
Standard errors in parentheses

Table 2.IA.4.6. Evaluating the robustness of the adapted functional regression model

The results confirm that for our simulated data (SD):

$$\beta_{SD} \cong \overline{RI}_{SD}^{DiD}$$

In fact, $\beta_{SD} = \overline{RI}_{SD}^{DiD} = -5$ exactly for the simulation results.

We note that the $\overline{RI}^{DiD}$ computed on the vertical stacks is exact in all cases of the simulation exercises because of the absence of measurement errors in the simulated treated and control samples and because, by construction, the mean anti-symmetry of the distribution of the time fixed effects around the event dates is also zero. If these conditions are not met, the relationship, though robust, may only be approximate. However, the main goal of checking that the functional regression model achieves a robust estimation is realized.

Chapter 3

Religiosity, Higher Purpose, and the Effectiveness of Intense Board Oversight

3.1. Introduction

Does the personal religiosity (or sense of higher purpose) of independent directors affect their approach to, or the results of their, intense board monitoring activities? This question is essential because Faleye, Hoitash, and Hoitash (2011—hereafter FHH) show that firms with monitoring-intensive boards exhibit a more effective board oversight, namely: greater sensitivity of CEO-turnover to [1-year] firm performance; lower excess CEO compensation; and reduced earnings management. However, they also show that oversight improvements obtained through intense board monitoring may come with lax engagement in board advisory activities, the combination of which ultimately leads to a net negative impact on firm value. Previous literature also argues that intense monitoring destroys trust and hampers communication between the chief executive officer (CEO) and the independent directors (Holmstrom, 2005; Adams and Ferreira, 2007); and reduces the amount of strategic information that the directors receive from management (Adams, 2009; Song and Thakor, 2006). Much of this friction, however, might relate to both the CEO's career (Milbourn, Shockley, and Thakor, 2001; Song and Thakor, 2006) and the independent directors' reputational concerns (Tepstra et al., 1993; Barnett et al., 1996; Mazar et al., 2008; and Longenecker et al., 2004). Suppose independent directors' religiosity is an attribute that can alter their approach to intense board monitoring. In that case, it becomes pertinent to explore whether such a change in perspective can produce results that are likely to ease the CEO-versus-independent-directors friction. If yes, then the expected benefits of intense board monitoring could accrue to the firm optimally.

We study directors' religiosity as a relevant attribute for two main sets of reasons. First, the literature in economics that examines the interlinkages or effects of religiosity and a sense of higher purpose on corporate behavior (and outcomes) is burgeoning. Religiosity and the pursuit of organizational higher purpose exhibit some fundamental parallels—both relate to beliefs in something bigger than oneself (e.g., see Lehrer and Chiswick, 1993; Weber, 1905; Starke and Finke, 2000; Lehrer, 1995, 2004; Stark and Finke, 2000; Barro and McCleary, 2003; Chen et al., 2016) or the pursuit of objectives that are far bigger than the rudimentary existential goals of an entity (e.g., see Thakor and Quinn, 2020). Their authentic internalization by agents (e.g., see Rossi, 2014) or permeation through an organization (e.g., see Serafeim and Gartenberg, 2016; Thakor and Quinn, 2020) leads to pro-social behavioral dispositions and, sometimes, transformational outcomes. Cranney (2013) shows that, compared to their non-religious counterparts, religious[54] individuals report a higher sense of purpose. Maxwell (2002), Wabara (2005), Ibarra (2015), Kim and Mauborgne (2015), and Quinn and Thakor (2019) all provide several anecdotes that suggest compelling links between founders' and leaders' senses of higher purpose and the "pursuits of higher purpose"[55] in the organizations that they lead. Also, in a recent survey of over 1,000 individuals, Bunderson and Thakor (2020) find a positive correlation between individual and organizational higher purpose.

Second, economic theories of moral behavior and identity suggest that people "care about who they are" and infer their values and preferences from past choices (Benabou

[54] Cranney (2013) refers to religious individuals as those who indicate that they are confident in God's existence.

[55] See Thakor and Quinn (2020) for theoretical details.

and Tirole, 2011) and that identity—a person's sense of self—affects how s/he behaves while interacting with others (Akerlof and Kranton, 2000). Furthermore, multiple[56] religious scriptures (Christian and non-Christian) identify and teach forbearance or patience as a vital tenet of the adherents' religiosities. Thakor and Quinn (2020), in their *"higher purpose, incentives, and economic performance"* theory suggest, among several results, that the adoption of a longer-term perspective and the delegated exercise of pro-social behavior are close to what they had in mind with *the pursuit of higher purpose*. They also suggest that when stakeholders care about a firm's organizational higher purpose, employees' equilibrium wages will tend to be lower. Cai, Kim, Lim, and Pan (2019—hereafter CKLP) uncover that firms with religious CEOs are associated with reduced earnings management. However, as FHH show, earnings management is not the sole responsibility of the CEO. This latter fact has been more prescient since August 2002, when the New York Stock Exchange (NYSE) required independent directors to meet at regularly scheduled executive sessions without management (e.g., see Browning and Sparks, 2015). Thus, independent directors (who, by definition, are not employees of the firm but an influential part of the corporate leadership at the board level) might, through their religiosity, contribute to the infusion and propagation of the expected gains of an authentic sense of higher purpose in the firm. We conjecture that any such benefit would show up as a significant difference in the effects of their intense board oversight.

Therefore, our goal in this chapter is twofold. On the one hand, we wish to empirically examine whether the causal effects of intense board monitoring uncovered by FHH vary depending on the independent directors' religiosity (or sense of higher purpose). On the other hand, we wish to broadly analyze how and to what extent these differences (if any) can help minimize the career and reputational frictions between CEOs

[56] For example, see https://en.wikipedia.org/wiki/Patience.

and the monitoring-intensive directors. It is natural then that our empirical strategy mainly builds on FHH's. Nevertheless, we proceed systematically in two primary analytical steps. In the first, we work mostly parallel to FHH: we collect financial and accounting data on S & P 1500 firms; we also collect similar biographical information on the firms' directors. We assemble our data from multiple sources—e.g., we gather much of the biographical data from BoardEx but complement this database with the ISS (formerly RiskMetrics) Directors Data. We corroborate or augment much of this information by hand-collecting supplementary biographical information from LinkedIn, Bloomberg, and other biographical websites. We collect the accounting and financial data on the firms from Compustat, the Center for Research in Security Prices (CRSP) database, Execucomp, and, generally, through the Wharton Research Databases Services (WRDS) platform. Overall, we collect more extensive data, and, notably, our sample length is more than double[57] FHH's.

Understandably, we begin our analyses by taking another look at monitoring-intensive boards and the benefits of intense board oversight. In other words, we essentially examine the evolution of the FHH results over our more extended sample period. This preliminary exercise also technically serves to validate our variables' construction. We also take the opportunity to perform some analyses that set the stage for our investigations in the second analytical step. For example, we further examine whether the difference in the CEO-turnover sensitivity to firm performance between monitoring-intensive and non-monitoring-intensive directors (that FHH uncover) relates to some intrinsic difference in perspective on the optimal evaluation horizon for CEO performance. Specifically, in

[57] Precisely, FHH used a data sample that spans the period from 1998 to 2006. In contrast, we use a base sample that spans the period from 1998 to 2018. In other words, we extend FHH's sample length by 12 years such that: (a) greater than 83% of our firm-years belong to the post-2006 period, and (b) a large chunk of our base sample belongs to periods that are much further from the introduction (in 2002) of the Sarbanes-Oxley Act (or SOX).

addition to evaluating the CEO-turnover sensitivity to 1-year firm performance (as did FHH), we reanalyze the same sensitivity (but) to a 2-year firm performance horizon. We adopt a robust definition of firm performance, precisely as the firm's stock return less same period return on the CRSP value-weighted portfolio of NYSE/Amex/Nasdaq stocks.

As we detail in the designated Internet Appendix, we find general consistencies with the main FHH results. In other words, monitoring-intensive boards continue to exhibit a more effective board oversight, namely: greater sensitivity of CEO-turnover to [1-year] firm performance; lower excess CEO compensation; and reduced earnings management. We find, however, that the difference in the CEO-turnover sensitivity to firm performance between monitoring-intensive and non-monitoring-intensive directors (that FHH uncover) does not relate to some intrinsic difference in perspective on the optimal evaluation horizon for CEO performance. The latter result is beneficial because it constitutes a useful reference for a similar investigation in our second analytical step. Hence, in a nutshell, compared to monitoring-intensive boards in general, non-monitoring-intensive boards do not preferentially hold a longer-term view of CEO performance. However, we find that they are by no means indifferent to long-term firm underperformance. We further find that while even in the event of forced turnover, it is significantly challenging to remove a CEO that doubles as the board's chairperson, CEO ownership matters far less in the event of a forced than in the event of voluntary turnover.

Next, in the second step, we return to the fundamental question posed in this chapter. We start by assuming that a good proxy for the individual religiosity (or sense of higher purpose) of the independent directors exists. We then adapt several FHH definitions and concepts to develop a novel two-dimensional measure[58]. Indeed, this

[58] See section 3.2.2 for a detailed description of how we construct our novel two-dimensional measure to sort firms based on their monitoring-intensive directors' aggregate religious inclination.

measure practically serves to sort the monitoring-intensive independent directors into two categories—i.e., the religious and non-religious intense board monitors. In other words, our measure allows us to identify a corporate board as either religious or non-religious monitoring-intensive in any given year. Subsequently, we develop three main conjectures[59] relating to this chapter's principal concern, systematically decompose our hypotheses into very simple and testable ideas, and then evaluate them sequentially. However, to put our novel measure into practical effect, we must first identify a good proxy for the independent directors' religiosity. We use[60] the religious affiliation of the academic institutions attended by the independent directors as a proxy for their religiosities. We consider the latter suitable for our empirical analyses because the noise in the proxy works against us finding an effect.

In all, we systematically formulate our empirical framework and testing strategies to eliminate (or drastically minimize) any potential endogeneity concerns. For example, FHH show that the benefits of intense board oversight are causal; hence, we mainly condition our empirical analyses of religiosity's marginal effects on boards being already monitoring-intensive. In other words, our empirical studies in this second analytical step principally evaluate whether there are differences in approach and benefits that arise from the aggregate religious inclinations of the monitoring-intensive directors, in particular. The intuition is also straightforward: as FHH show, monitoring-intensive directors, be they religious or not, will have a specific family of impacts on a board's oversight quality (relative to their non-monitoring-intensive counterparts). However, drawing from the literature, we anticipate that the strengths of the (monitoring-intensive directors') impacts

[59] See section 3.2.1 for the hypotheses.

[60] We are not the first to use a version of this proxy for individual religiosity. CKLP used the undergraduate educational experience in church-affiliated colleges as a proxy for CEOs' religiosity. However, we use a slightly different variant. See section 3.2.3 for a detailed analysis and justification of our version of this proxy.

might vary depending on their religiosity. Therefore, our marginal analyses in this second analytical step are technically equivalent to examining whether the known (i.e., published) causal effects of a two-category variable differ significantly between the categories.

We find that, compared to their non-religious counterparts, religious monitoring-intensive directors exhibit significantly lower sensitivity of CEO turnover to firm performance over a holding period of 1 year. However, for the more extended holding period of 2 years, this difference in sensitivity significantly switches direction, consistent with the "*higher purpose, incentives, and economic performance*" theory by Thakor and Quinn (2020), suggesting that believers in higher purpose will tend to hold a longer-term perspective. We also find that religious monitoring-intensive directors further significantly reduce earnings management, on average, all else equal. Similar evidence that they tend to reduce excess total CEO compensation further, though ordinarily not as strong, becomes highly significant when the lead independent director and/or a majority of the principal monitoring committee chairs are also religious, consistent with Wabara (2021a). We also find that the monitoring-intensive directors' religiosity is more economically significant than the CEOs' for the quality of the earnings information in firms' financial reports.

Finally, we also find that the differences in average sensitivities of CEO turnover to firm performance do not take effect, statistically, until individual firm performances worsen beyond -5% relative to the market, at the minimum. This result is unchanged either for monitoring-intensive versus non-monitoring-intensive boards or for religious monitoring-intensive versus non-religious monitoring-intensive boards.

A residual endogeneity concern that might linger at this stage relates to whether our results are associated with some fundamentally "conservative" firms; after all, per the literature, conservative firms typically possess an integrity culture. For example, one might imagine that the religious independent directors systematically select to serve as directors only in conservative firms. However, the nature of (and the extent of the within-firm variations in our novel two-dimensional measure for the intense board monitors' aggregate religiosity) suggests that a selection story could not explain our results. Notwithstanding, we do more than basic rationalizations. We effectively use the flexibility inherent in our measure's two-dimensionality to rule out this residual reverse causality concern. We detail the empirical analyses in section 3.7. The underlying intuition is also straightforward: thus far, our results indicate that the independent directors' aggregate religiosity takes effect mainly when they are monitoring-intensive. Now, suppose not (i.e., assume, temporarily, that our results are simply due to the religious directors' preselecting to serve as directors only in firms with an integrity culture). In that case, we should obtain similar results when comparing the religious and non-religious but non-intense board monitors. We do not.

This chapter makes several contributions. First, we extend the literature on boards' oversight duties by providing evidence that the religiosity (or the sense of higher purpose) of independent directors affects their approach to and the overall benefits of their intense monitoring activities. Our contribution to this literature is significant because although FHH show that firms with monitoring-intensive boards exhibit more effective board oversight, namely: greater sensitivity of CEO-turnover to [1-year] firm performance; lower excess CEO compensation; and reduced earnings management, they also show that some of these benefits (e.g., greater sensitivity of CEO-turnover to [1-year] firm performance) come at high costs that ultimately impact the value of the firm, negatively. To this end, Holmstrom (2005) argues that intense monitoring destroys the trust necessary for the CEO to share relevant strategic information with the directors, and Adams (2009) provides survey evidence in support. However, we show that religious directors further reduce earnings management and excess CEO compensation but differentially exhibit lower (greater) sensitivity of CEO turnover to short (long) term firm performance when spearheading the board's intense monitoring. This result is crucial for at least two reasons: First, religious monitoring-intensive directors' lower (or almost non-existent) sensitivity of CEO turnover to 1-year firm performance is likely to help minimize CEOs' career concerns (e.g., see Milbourn, Shockley, and Thakor, 2001; Song and Thakor, 2006), preserve trust, facilitate better relationships, enhance strategic information sharing and thereby improve the dynamics of the board (e.g., see Adams and Ferreira, 2007). Second, their more heightened sensitivity to (and possible disdain for) longer-term firm underperformance, however, is likely to help preserve long-term shareholder value.

We further contribute to the literature on boards' oversight duties by providing additional empirical evidence that can become established in the family of stylized facts in this area. For example, our findings that (i) although when compared to monitoring-intensive boards in general, non-monitoring-intensive boards do not preferentially hold a longer-term view of CEO performance, they are by no means indifferent to long-term firm underperformance, (ii) while in the event of forced turnover, it is significantly challenging to force a CEO that doubles as the chair of the board out, CEO ownership matters far less in the event of a forced than in the event of voluntary turnover, and (iii) whether in consideration of monitoring-intensive/non-monitoring-intensive boards or religious monitoring-intensive/non-religious monitoring-intensive boards, the differences in average sensitivities of CEO turnover to firm performance do not take effect, statistically, until firm performance worsens beyond a minimum of -5% relative to the market.

We also contribute to the burgeoning literature in economics that examines the effects of religiosity and a sense of higher purpose on corporate behavior and outcomes. Indeed, our work is related to CKLP, which suggest that religious CEOs contribute to setting the tone at the top by improving earnings quality. However, referring to the causal impacts of monitoring-intensive independent directors uncovered by FHH, we note that reducing earnings management is not the CEO's sole responsibility. We show that the effect of independent directors' religiosity on improving earnings quality might be more economically significant. We also connect independent directors' religiosity to the infusion or propagation of an authentic sense of higher purpose in the firm: Cranney (2013) shows that, compared to non-religious people, religious individuals report a higher sense of purpose. Thakor and Quinn (2020) suggest that a longer-term perspective is integral to pursuing an organizational higher purpose. They also suggest that when the corporate higher purpose is authentic, and stakeholders care about it, the equilibrium wages (of all employees) are lower. In turn, we show that relative to their non-religious counterparts, religious monitoring-intensive directors exhibit significantly lower sensitivity of CEO turnover to firm performance over a holding period of 1 year. However, for the longer holding period of 2 years, this difference in sensitivity significantly switches direction. We also show that religious monitoring-intensive directors further reduce excess total CEO compensation, especially when the lead independent director or a majority of the principal monitoring committee chairs are also religious. Overall, our findings suggest that religiosity could be a source of an authentic sense of organizational higher purpose.

The rest of this chapter continues as follows. Section 3.2 provides background information. It includes our development of the principal hypotheses for this study, how we build the novel two-dimensional measure to sort firms based on their monitoring-intensive directors' aggregate religious inclination, and our choice of a good proxy for the directors' religiosity. Section 3.3 presents our data, variables, and descriptive statistics. Sections 3.4, 3.5, and 3.6 discuss our results on the effects of independent directors' religiosity on CEO turnover, CEO compensation, and firms' earnings quality, respectively. Section 3.7 outlines our robustness tests and other related analyses. Section 3.8 concludes.

3.2. Background Information

This section outlines the development of our main hypotheses for this chapter. It describes how we build our novel two-dimensional measure to sort firms based on their monitoring-intensive directors' aggregate religious inclination. It also presents and justifies our choice of proxy for the personal religiosity of the independent directors.

3.2.1. Hypotheses development

A goal of ours is to build on the FHH results. Specifically, suppose we can create a device to effectively sort directors, particularly the monitoring-intensive independent directors, into two categories—i.e., religious and non-religious. In that case, we could evaluate any differences in their respective impacts on oversight results relative to their non-intense counterparts. Consequently, our first line of inquiry relates to how independent directors' religiosity might affect the sensitivity of CEO turnover to firm performance, as uncovered by FHH. We note that factors[61] other than firm performance could impact CEO turnover. Nevertheless, at this stage, we begin by adopting the robust version of firm performance (as defined by FHH) for direct comparative causal analysis.

In other words, we anticipate using the same sensitivity of CEO turnover to firm performance variables construction strategy (and empirical controls) as did FHH while

[61] For example, Huang, Maharjan, and Thakor (2020) propose and test a new explanation for forced CEO turnover. They suggest that investors may disagree with management on optimal decisions due to heterogeneous prior beliefs (about project choices). They provide evidence that because such disagreement may be persistent and costly, firms sometimes resort to replacing CEOs with whom investors disagree. They further document that, in other instances (and after controlling for firm performance), a lower level of CEO-investor disagreement may serve to partially "protect" CEOs from being fired, thus reducing turnover-performance sensitivity.

simultaneously accounting for CEOs' religiosity. To this end, however, we draw intuition from a wide array of literature. First, we note that multiple religious scriptures (Christian and non-Christian) identify and teach forbearance or patience as a vital tenet of the adherents' religiosities. We also acknowledge the essential facts emanating from the burgeoning literature on the effects of religiosity: In economic theories of moral behavior and identity, people "care about who they are" and infer their values and preferences from past choices (e.g., see Benabou and Tirole, 2011); and identity—a person's sense of self—affects how s/he behaves while interacting with others (e.g., see Akerlof and Kranton, 2000). Cranney (2013) shows that, compared to non-religious people, religious individuals report a higher sense of purpose. Maxwell (2002), Wabara (2005), Ibarra (2015), Kim and Mauborgne (2015), and Quinn and Thakor (2019) all provide several anecdotes that suggest compelling links between founders' and leaders' senses of higher purpose and the pursuits of higher purpose in the organizations that they lead. In a recent survey of over 1,000 individuals, Bunderson and Thakor (2020) find a positive correlation between individual and organizational higher purpose. Thakor and Quinn (2020), in their *"higher purpose, incentives, and economic performance"* theory, also suggest that the adoption of a longer-term perspective and the delegated exercise of pro-social behavior are close to what they had in mind with *the pursuit of higher purpose*.

Therefore, the religious independent directors might forbear more, possess a higher sense of purpose, and be more intrinsically inclined to have a longer-term perspective or take a more patient approach to CEOs' performance evaluation. It follows then that the sensitivity of CEO turnover to firm performance between the religious and the non-religious independent directors could vary with the CEO performance evaluation horizon. For instance, in the shorter-term, the non-religious independent directors could exhibit a greater sensitivity of CEO turnover to firm performance. However, over a more extended performance measurement horizon, this difference in sensitivity may significantly switch directions or, at least, statistically dissipate. Consequently, we hypothesize as follows.

Hypothesis 1: *Compared to their non-religious counterparts, religious monitoring-intensive directors, on average, exhibit a lower sensitivity of CEO turnover to firm performance over a holding period of 1 year. However, for a holding period of 2 years, the difference in sensitivity is likely to significantly switch direction or statistically dissipate, all other things equal.*

Next, we turn to our second line of inquiry about whether non-religious directors may have a higher propensity to overcompensate the Chief Executive Officer. Thakor and Quinn (2020), in their *"higher purpose, incentives, and economic performance"* theory, further suggest that when stakeholders care about a firm's organizational higher purpose, the equilibrium wages of employees (of which the CEO is one) will tend to be lower. Previous research also suggests that religious people may hold more conservative moral standards than their non-religious counterparts and that personal religious commitments may be significantly associated with higher levels of ethics in business (e.g., see Tepstra et al., 1993; Barnett et al., 1996; Mazar et al., 2008; and Longenecker et al., 2004). Furthermore, identity—a person's sense of self (e.g., ethical, fair, etc.)—affects how s/he behaves while interacting with others (e.g., see Akerlof and Kranton, 2000). As such, relatively, and to the extent[62] that stakeholders care about a firm's organizational higher purpose or ethical behavior correlates with a sense of fairness (generally or in CEO compensation), we would expect intense board oversights driven by religious independent directors to be associated with less excess CEO compensation, particularly so concerning

[62] We also note that Smith, Ryan, and Diggins (1972) and Michaels and Miethe (1989) find no effect of religion on academic integrity. Heatherington and Feldman (1964) even suggest a positive link between religion and cheating.

popular pay items such as cash and equity. Conversely, we would also expect intense board oversights spearheaded by non-religious independent directors to be associated with significantly greater excess total CEO compensation. Thus, we hypothesize as follows.

Hypothesis 2: *Compared to their non-religious counterparts, religious monitoring-intensive directors are associated with significantly lower excess CEO compensation, particularly so concerning popular pay items such as cash and equity, ceteris paribus*.

Finally, we turn to our third line of inquiry about whether non-religious directors may have a higher likelihood of being associated with firms that present with lower earnings quality. Previous studies allude to a positive association between religiosity and risk-aversion (e.g., see Miller and Hoffman, 1995; Barsky et al., 1997; Diaz, 2000; Miller, 2000; Osoba, 2003; CKLP). Past studies also suggest a positive association between risk-aversion and reputational concerns (e.g., see Hughes and Thakor, 1992; Boquist, Milbourn, and Thakor, 2010). Therefore, to the extent[63] that religious independent directors are likely to be individually more risk-averse than their non-religious counterparts, we expect intense board oversights dominated by the religious independent directors to be associated with reduced[64] earnings management. Altogether, the greater relative risk-aversion and the long-term perspective of the religious directors might imply a stronger preference to distance themselves from future financial restatements (and perhaps other public repercussions) that might adversely affect their reputations as corporate board directors or their board careers in the long term. Conversely, we would also expect intense board oversights spearheaded by non-religious independent directors to be associated with lower earnings quality. Thus, we further hypothesize as follows.

[63] Goel and Thakor (2003) develop a model in which earnings smoothing is motivated by the desire to reduce the perceived volatility of firms' earnings.

[64] Risk-aversion could also lead to information concealment to save face.

Hypothesis 3: *Compared to their non-religious counterparts, religious monitoring-intensive directors are more likely to further reduce earnings management, all other things equal.*

3.2.2. Sorting monitoring-intensive boards by religiosity

First, we assume that there exists a good proxy for the religiosity of the independent directors. Next, using the FHH definitions, we identify the monitoring-intensive independent directors and then reconstruct the FHH variable for a monitoring-intensive board—hereafter, the FHH measure (or FHHm for short). FHH define a monitoring-intensive director as one that serves in at least two of the three principal monitoring committees and define a monitoring-intensive board for any given year as one in which a majority of the independent directors are monitoring-intensive. The principal monitoring committees are the audit, the compensation, and the nominating/governance committees.

In other words, $FHHm_{i,t}$ is a dummy variable that takes the value of 1 for firm i, in year t, if the board meets the FHH definitions for intense board monitoring:

$$FHHm_{i,t} = \begin{cases} 1, & Monitoring - intensive\ board \\ 0, & Non - monitoring - intensive\ board \end{cases}$$

However, to sort firms by their monitoring-intensive directors' aggregate religious inclination, we develop a novel two-dimensional measure, as schematized in Figures 2 and 3. Specifically, to determine which type (by religiosity) of independent directors is mostly responsible for the intense monitoring on a board in a given year, we first compute:

$$\mu_{i,t}^{R_g} = [\langle Num_{MC}^{RI} - Num_{B}^{RI} \rangle - \langle Num_{MC}^{NI} - Num_{B}^{NI} \rangle]_{i,t}$$

$$\mu_{i,t}^{R_g} : \begin{cases} < 0 \\ = 0 \\ > 0 \end{cases}$$

Where $\mu_{i,t}^{R_g}$ is a religion-intense monitoring parameter for board i in year t that takes a negative (positive) value when the intense monitors on the board are mostly non-religious (religious) but equals zero when the intense monitoring on the board is equally driven by both the non-religious and religious independent directors (i.e., the religion-indifferent intense board monitoring case) or when the monitoring on the board is simply non-monitoring-intensive. Num_B^α and Num_{MC}^α are the composite[65] numbers of the type- α independent directors on the board and the monitoring committees, respectively. $\alpha = RI$ and $\alpha = NI$ signify religious-independent and non-religious-independent, respectively.

To allow for the possibility of a coarse aggregation of our firm-period religion-intense monitoring parameters, we define a generalized version as follows:

$$MWm_{i,t,n} = \frac{\sum_{\tau=0}^{n} \omega_{t-\tau}\mu_{i,t-\tau}^{R_g}}{n+1}$$

Where, for ease of reference, we call $MWm_{i,t,n}$ the Milbourn-Wabara measure for the religious inclination of the intense board monitors for firm i, in period t, aggregated over n preceding periods up to period t. n is an integer that ideally may be less than 2 but could also depend on some desired technical meaning, the level of coarseness sought, or the minimum number of firm-periods per firm contained within the entire sample. $\omega_{t-\tau}$ is a weighting factor that may depend on the importance of the $(t-\tau)th$ period to the period-t event. If $n = 0$, and $\omega_t = 1$ for all t, then $MWm_{i,t,0} = \mu_{i,t}^{R_g}$ and is strictly equivalent to a firm-period sample. For simplicity and without loss of generality we, in this study, set $n = 0$ and $\omega_t = 1$ for all t, where t is in years.

[65] For example, if there are only eight independent directors on a corporate board but precisely two of them each serve in exactly two different monitoring committees while the rest each serve in one monitoring committee, then, clearly, $Num_B = 8$, while $Num_{MC} = 8 + 1 + 1 = 10$, for that board. We follow the same logic to compute Num_{MC}^{RI}, Num_B^{RI}, Num_{MC}^{NI}, and Num_B^{NI}; with RI signifying religious-independent and NI signifying non-religious-independent. Again, see Figures 3.1 and 3.2 for the schematic descriptions.

To further simplify our empirical comparisons, we take a marginal approach to constructing the dummy variable for our two-dimensional measure for the religious inclination of the intense board monitors for firm i, in year t. Specifically, we set:

$$MWm_{i,t} = \begin{cases} 0, & MWm_{i,t,0} = \mu_{i,t}^{R_g} < 0,\ FHHm_{i,t} = 1 \\ 1, & MWm_{i,t,0} = \mu_{i,t}^{R_g} > 0,\ FHHm_{i,t} = 1 \end{cases}$$

This duo-conditional rendering of our measure states that a corporate board for firm i is non-religious monitoring-intensive or a non-religious intense board monitor in year t (i.e., $MWm_{i,t} = 0$) if the board is monitoring-intensive (i.e., $FHHm_{i,t} = 1$) and a higher number of the straddling across the three principal monitoring committees is spearheaded by the non-religious independent directors (i.e., $MWm_{i,t,0} = \mu_{i,t}^{R_g} < 0$). Conversely, it also states that a corporate board for firm i is religious monitoring-intensive or a religious intense board monitor in year t (i.e., $MWm_{i,t} = 1$) if the board is monitoring-intensive (i.e., $FHHm_{i,t} = 1$) and a higher number of the straddling across the three principal monitoring committees is spearheaded by the religious independent directors (i.e., $MWm_{i,t,0} = \mu_{i,t}^{R_g} > 0$). Notice that we exclude all cases for which $\mu_{i,t}^{R_g} = 0$ or $FHHm_{i,t} = 0$ (i.e., we exclude the religion-indifferent monitoring-intensive cases and the non-monitoring-intensive cases—both the passive[66] and the non-passive components).

3.2.3. Proxy for the personal religiosity of the independent directors

To put our novel device (for sorting firms based on their monitoring-intensive directors' aggregate religious inclination) into practical effect, we ultimately find a good

[66] By a "passive" non-monitoring-intensive case, we mean that each of the independent directors on the board in the given year serves in, at most, one monitoring committee.

proxy for the independent directors' religiosity. We emphasize that this device does not depend on any particular proxy. Instead, it takes a good proxy for personal religiosity and sorts the independent directors into two main categories—religious and non-religious.

We use the religious affiliation of the academic institutions attended by the independent directors as a proxy for their religiosities. Indeed, we are not the first to use a variant of this proxy for individual religiosity. CKLP used the undergraduate educational experiences in church-affiliated colleges to proxy for CEOs' religiosities. Nevertheless, we notice that many directors who hold a graduate degree in Divinity are also religious priests. We further observe that many graduate academic institutions with religious affiliations document or avow the infusion of religiosity in the educational content and other aspects of their graduate programs. Hence, we refine this proxy by additionally flagging a director or a CEO as potentially religious if their graduate degree is in Divinity or from a religious-affiliated graduate program that explicitly declares the infusion of religious morality in the study of business or law. We consider the resultant proxy particularly good for our empirical analyses for several reasons.

First, Sacerdote and Glaser (2008) explain the substitution relationship between education and religiosity, documenting that the more educated people are, the less the fervor of their religious beliefs. Therefore, the affiliations of undergraduate and graduate programs to religious organizations and the explicit assertions by the programs to infuse the tenets of their religiosity into students' learning experiences would counteract the diminishing impact of higher education on the attendees' prior religiosity. In effect, religious affiliations will tend to shift the typical relationship between education and religiosity from substitution to complementarity. A couple of things could then follow: people who fear the potential adverse impact of higher education on their religiosity might systematically choose to attend religious-affiliated educational institutions. Others with no prior religiosity who attend religious-affiliated institutions might become religious, partially at least. The proxy may also be noisy: non-religious individuals might attend religious-affiliated academic institutions and remain non-religious, while some religious individuals might go to secular institutions and still retain their religiosity. Practically, other shades of possibilities also exist. However, as long as activating our sorting device with this proxy is not systematically related to the set of corporate outcomes we examine (i.e., the sensitivity of CEO turnover to 1 or 2-year firm performance, CEO compensation, and earnings management), the noise in the proxy works against us finding statistically significant effects of independent directors' religiosity on board oversight quality.

Second, a set of consistent facts in the literature provide strong support for our judgment of this proxy's appropriateness for directors' or CEOs' religiosity: CKLP show that this proxy is robust to controlling for school rankings, omitted firm characteristics, CFO religiosity, geographical location, and that earnings management is lower for firms managed by CEOs who attended protestant colleges. Barsky, et al. (1997) and Shu, et al. (2007) show that protestants are relatively more risk-averse than other religious groups. Sacerdote and Glaeser (2008) show that religious beliefs are stronger among protestants.

3.3. Data, Sample, and Variables

Figure 3.1 schematizes the generation of our base sample. Similar to FHH and CKLP, our sample consists of S & P 1500 firms. We obtain data on the firms from several sources. Data on board attributes come mainly from the BoardEx database. We also complement this database with the ISS (formerly RiskMetrics) Directors Data. Both databases provide detailed information on each director, including such items as age, gender, primary occupation, independence status, and service on the three principal monitoring committees. The BoardEx database, however, provides other biographic information, such as the educational backgrounds of the directors, not available from the ISS Directors Data. Our data from these databases span the periods from 1997 to 2019 and 1998 to 2019, respectively. Different from FHH, however, we are primarily interested in, not only the independent status of the directors and their service on the three principal committees but also their educational background and religious affiliations. Moreover, whereas we similarly start our base sample from 1998, we extend it beyond 2006 (as in FHH) to 2018.

We obtain data on the religious affiliations of the academic institutions attended by the directors from the U.S. Department of Education. We obtain accounting data from Compustat, stock return data from the Center for Research in Security Prices (CRSP) database, CEO compensation data from Execucomp, age and all turnover data from BoardEx and Execucomp, and forced turnover data from the contributed[67] data on the Wharton Research Databases Services (WRDS) website. We begin the construction of our variables for board characteristics in BoardEx. After eliminating delisted firms, observations with unknown annual report dates, and firm-years in which the educational backgrounds and/or the religious affiliations thereof of all directors are not fully[68] known (or we are unable to hand-collect the missing data), we retain a total of 26,382 firm-years.

We also eliminate firms from sectors such as banking and utilities, owing to the significant differences in regulatory oversight that can limit the role of a board, thus further reducing our sample by 2,854 firm-years. Consequently, our base sample (i.e., the compact block of data before the respective merges with an array of financial and accounting data, or before the final construction of the Milbourn-Wabara measures for

[67] See Peters and Wagner (2014).
[68] Interestingly, the loss of a few firm-year(s) due to any missing religiosity data had zero net impact on our core analyses dataset (i.e., dataset post-intersection with the array of financial data).

this study) includes 23,528 firm-years[69] of board characteristics, for 2,945 unique firms, from 1998 to 2018. Using this base sample, we construct the indicator variables for the Milbourn-Wabara measures, as detailed in section 3.2.2. We, below, further discuss these variables and, in later sections, discuss other variables as they appear in the specific tests. See Table 3.17 for the full list and detailed definitions of over forty variables used in our regression models.

3.3.1. Variables construction

We begin our variables construction process by intersecting the educational institutions attended by all directors in the BoardEx database with the religious affiliation data from the U.S. Department of education. Similar to CKLP, we flag as religious all directors with undergraduate degrees from academic institutions affiliated with churches such as Baptist, Roman Catholic, United Methodist, and so forth. Different from CKLP, however, we also flag as religious any director with a graduate degree in Divinity or a graduate degree in law and/or business from a church-affiliated institution that explicitly declares the infusion[70] of religious morality as a fundamental goal of the program. To stay consistent, we keep 100% of the firms (or more precisely, firm-years) for which the educational background and religious affiliation[71] data are available for all directors listed

---

[69] Actual sample size for each regression depends on the minimum sample length of the main and control variables.

[70] We observe that many directors with graduate degrees in divinity are religious priests. Also, many church-affiliated graduate programs avow the infusion of religious values in their teachings. For example, in their ranking of the 50-Best Value Christian MBA programs, the Christian Universities Online (CUO) writes: "Employers continue to target MBA graduates when seeking new hires. The institutions in our ranking not only position their graduates for a significant increase in their salary, but they also equip and train their graduates to integrate their Christian faith into the fast-paced and challenging environment they will face upon graduation." Attendance also captures pre-college religiosity. See https://www.christianuniversitiesonline.org/best-value-christian-mba-programs/

[71] As further discussed in Section 3.6, some measurement errors might be embedded in our proxies and/or measures. For instance, it is indeed possible that, on the one hand, directors who are non-religious might have attended church-affiliated academic institutions; and on the other hand, religious directors might have attended secular academic institutions. However, and as similarly highlighted in CKLP, to the extent that measurement errors are not systematically and homogenously associated with the respective parameters for

on the board in that year. Nevertheless, and whenever possible, we save a few firm-year samples (typically for an entire firm) by hand-collecting from the internet[72], the biographical data of directors whose educational backgrounds are unavailable in BoardEx.

Following FHH, we identify each independent director as either monitoring-intensive (if the director serves in at least two principal monitoring committees) or not (if otherwise). We also identify a board as either monitoring-intensive in any given firm-year (if a majority of the independent directors are monitoring-intensive) or not (if otherwise). However, to further conceptualize how we capture the religiosity of the monitoring-intensive boards, consider that each monitoring-intensive director is first assigned to an initial principal monitoring committee, which we call a director's base committee. FHH define a director as becoming monitoring-intensive when s/he begins to participate in the activities of other principal monitoring committees. We call this additional participation in the activities of other monitoring committees a straddling from the base committee to that (those) extra committee(s). We are essentially indifferent to the exact identity of the base committees, but are particularly interested in the number of straddles per director, which signify increased attention of directors to board monitoring activities.

Since the effectiveness of board-oversight is increasing in the number of these straddles (e.g., see Vafeas 2005, Fich and Shivdasani, 2006; Ferris, Jaganathan, and Pritchard, 2003; FHH), we define a religious monitoring-intensive board or a religious intense board monitor as one in which a majority of the total straddling from a base (or initial) principal monitoring committee to another principal monitoring committee is done by the religious independent directors. We say the opposite when, on a monitoring-

CEO turnover, earnings management, and excess CEO compensation, the noisiness of our proxies/measures work against the general consistency we find in our results (and not otherwise). Moreso, and as emphasized in our rationale for using this proxy for directors' religiosities (e.g., see section 3.2.3), attendance in church-affiliated institutions captures not only the pre-college religiosity but also the avowed infusion of religious values into students' educational experience.

[72] LinkedIn, Bloomberg, et cetera. However, success with this effort was limited to just about two instances.

intensive board, a majority of the total straddling is done by the non-religious independent directors. We capture both of these definitions using a single indicator variable.[73]

We also control for the directors' external time commitment. To do this, we create an indicator variable for an externally busy board, which we define, similar to FHH and consistent with Fich and Shivdasani (2006), as one in which a majority of the independent directors serve on at least three corporate boards. We similarly create other control variables that affect directors' effectiveness including, among others, board size (which we also define as the natural log of the number of directors), board composition (the fraction of independent directors), and firm size (the natural log of market capitalization). Different from FHH, however, but complementary to CKLP, we systematically include an indicator variable for the religiosity of the CEO in our vector of controls. Again, see Table 3.17 for the full list and detailed definitions of the variables used in the various regression models.

3.3.2. Descriptive statistics

Table 3.1 presents the firm-year distribution of our base sample, by BoardEx sectors. There are a total of 37 sectors, of which none overly dominates the entire sample. Of the 23,528 firm-years, three sectors come the closest to 10%, but the remaining sectors have between 0.4% and 6.9%, each. Table 3.2 presents the annual distribution of the principal monitoring committees and the Milbourn-Wabara Measures. Each of the nine years before (and including) 2006 has less than 1000 firm-year data points for a total of 3,989 firm-years, amounting to less than 17% of our base sample. Conversely, each of the twelve years beyond 2006 has greater than 1000 firm-year data points for a total of 19,539 firm-years, amounting to over 83% of our base sample. Virtually every firm in our base

[73] See Figures 3.1 and 3.2 for schematic descriptions. See also Table 3.17 for more on the definitions of variables.

sample has an audit committee, 93% to 100% of the firms in our sample have a compensation committee, and 83% to 100% of the firms in our sample have the nominating/governance committee, except for the period between 1999 and 2003. This latter fact is because the nominating/ governance committee did not become common until 2004 (e.g., see FHH). Notably, the period from 1999 to 2003 constitutes less than 6.8% of our base sample. On average, across all the years in our base sample (i.e., from 1998 to 2018), 58% of the boards are monitoring-intensive[74], 90% are non-religious, and 10% are either religion-indifferent or religious.

Table 3.3 presents the descriptive statistics for some of the board and firm characteristics (after intersecting the base sample with an array of accounting and financial data). The average board has seven members, six of whom are independent. The median audit, compensation, and nominating/governance committees are virtually independent and, on average, each principal monitoring committee has four members. The boards in our sample are hardly externally busy. Our sample firms are also fairly large, with an average and median market capitalizations of $4.40 billion and $0.74 billion, respectively. The average and median total assets are $5.40 and $0.70 billion, respectively. Between 1998 and 2018, the median firm earned roughly a 10% annual return on assets (ROA). The median CEO is close to 56 years of age, owns about 0.4% of the firm, and receives total compensation of about $8.3 million. 40% of the CEOs also chair the board, on average.

[74] This is consistent with FHH, in whose sample the overall percentage of monitoring-intensive boards was 57%.

3.4. Religiosity, Firm Performance, and CEO Turnover

In this section, we test the first of our hypotheses on the marginal returns of religiosity (or a sense of higher purpose) to intense board oversight. Specifically, we evaluate whether compared to their non-religious counterparts, religious intense board monitors, on average, exhibit a lower sensitivity of CEO turnover to firm performance over a holding period of 1 year; and whether, for a holding period of 2 years, the difference in sensitivity significantly switches direction or statistically dissipates, all other things equal. Our analyses are principally marginal to the finding that intense monitoring is associated with better board oversight (e.g., see Weisbach, 1988; Yermack, 1996; Hermalin, 2005; FHH). Whereas FHH show that, compared to their non-monitoring-intensive counterpart, monitoring-intensive boards exhibit a greater sensitivity of CEO turnover to firm performance, we seek to evaluate whether and how, conditional on all boards being monitoring-intensive, religiosity (or a sense of higher purpose) also matters. Empirically, we adapt the FHH definitions for monitoring-intensive boards to develop a novel measure for sorting intense board monitors into religious and non-religious cohorts and refine a proxy for personal religiosity to activate our sorting device (e.g., see section 3.2).

For ease of reference, we call the generalized form of this device the Milbourn-Wabara measure. We also adopt the robust version of FHH's definitions of firm performance (i.e., as market-adjusted stock returns, where the market is defined as the CRSP value-weighted portfolio of the NYSE/Amex/Nasdaq stocks). We analyze and cross-validate all CEO transitions in both the BoardEx and the ExecuComp databases. We also analyze and cross-validate the contributed data on forced turnovers available on the WRDS website (e.g., see Peters and Wagner, 2014). We further control for factors—such as CEO duality, board independence, the external busyness of the directors, the size of the board, the ownership stake of the institutional investors, the fraction of the company owned by the CEO, the size of the firm, the age of the CEO—that can affect the replacement of a CEO (e.g., see Weisbach, 1988; Denis, Denis, and Sarin, 1997; Goyal and Park, 2002; FHH).

In contrast, however, while FHH used a data sample that spans the period from 1998 to 2006, our base sample runs from 1998 to 2018 (i.e., some additional 12 years; and much further from the introduction of the Sarbanes-Oxley Act, or SOX, in 2002) and over 83% of our firm-years come from the period post-2006. Also, whereas FHH composed their performance measures based only on a 1-year holding period, we seek to evaluate the progression of the sensitivity of CEO turnover to firm performance as the performance measurement horizon shifts to a term longer than one year. Consequently, we proceed in two main steps. First, we look at monitoring-intensive boards, firm performance horizons, and CEO turnover; and second, we look at religious-monitoring-intensive boards, firm performance horizons, and CEO turnover. We discuss these steps in detail below.

3.4.1. Monitoring-intensive boards, performance horizons, and CEO turnover

We take advantage of our longer sample data (spanning the 21 years from 1998 to 2018) to examine the stability and/or evolution of the related FHH results[75]. Tables A.1 and A.2/Figures A.1 and A.2 (in Appendix 3.IA and accompanied by granular details of the results) present our logistic regression model for this purpose/plot the predictive margins[76], respectively. In summary, we affirm that monitoring-intensive boards indeed exhibit a greater sensitivity of CEO turnover to firm performance, particularly when the performance measure is over a 12-month or 1-year horizon. Precisely, while the non-

[75] On this particular subject, FHH find that, compared to non-monitoring-intensive boards, monitoring-intensive boards exhibit a greater sensitivity of CEO turnover to [1-year] firm performance.

[76] Predictive margins are statistics calculated from predictions of a previously fit model at fixed values of some covariates and averaging or otherwise integrating over the remaining covariates (e.g., see Buis, 2010; Hosmer, Lemeshow, and Sturdivant, 2013; Long and Freese, 2014; Norton, Wang, and Ai, 2004; Williams, 2012).

monitoring-intensive boards appear steadily disinclined to force the CEO out even for a single-year market underperformance up to -20%, the probability of forced turnover rises exponentially for the monitoring-intensive boards. However, the picture changes slightly when the firm-performance measure is over a 24-month or 2-year horizon. For this longer-term horizon, while average sensitivities remain qualitatively higher for the monitoring-intensive boards, average sensitivities also increase for the non-monitoring-intensive boards (albeit more gently but) sufficiently high to preclude the divergences of the average sensitivities of the two board-types from being statistically significant. This suggests that, although when compared to monitoring-intensive boards non-monitoring-intensive boards do not preferentially hold a longer-term view of CEO performance, they are not systematically indifferent to long-term firm-underperformance.

3.4.2. Religious-monitoring-intensive boards, performance horizons, and CEO turnover

Next, we proceed to evaluate whether there exist any marginal effects of religiosity on these sensitivities. We condition our analyses on all boards being monitoring intensive. Effectively, we keep only the firm-years in which all boards are monitoring-intensive (i.e., 58% of our base sample). Next, we use the Milbourn-Wabara measure to sort these boards into religious and non-religious cohorts. Empirically, we run similar regressions to Tables A.1 and A.2, with the only difference being that the monitoring-intensive board variable is replaced with the religious monitoring-intensive board variable. The latter equals one when, on a monitoring-intensive board, a majority of the straddling from a base (or initial) principal monitoring committee to another principal monitoring committee is done by the religious independent directors, and zero when, on a monitoring-intensive board, a majority of the said straddling is done by the non-religious independent directors.

Table 3.4 presents the associated logistic regression model when the performance measurement horizon is one year. The dependent variable equals one for forced turnovers and zero otherwise (in columns 1 and 2) and equals one for all turnovers and zero otherwise (in columns 3 and 4). Similarly, our specific variable of interest at this point is the interaction term between religious monitoring intensity and 1-year performance, and consistent with our hypothesis, we find a positive coefficient, which is significant at the 1% level. We also compute the predictive margins for both the religious monitoring-intensive and the non-religious monitoring-intensive boards. Figure 3.4 graphs these predictive margins. We observe that over the 1-year holding period, both board-types, in this case, do not appear to be statistically inclined to force a CEO out for matching or for barely underperforming the market up to -5%. Below this level of market-underperformance, however, we find a significant statistical divergence. Specifically, while the religious monitoring-intensive boards remain steadily disinclined to force the CEO out after only a single period of underperformance, the probability of forced turnover rises exponentially for the non-religious monitoring-intensive boards, from 0 at 0%, to close to 0.6 at -20% (with a 95% CI between a little greater than 0.3 and close to 0.9).

Table 3.5 presents the associated logistic regression model when the performance measure horizon is 24 months or two years. The dependent variable equals one for forced turnovers and zero otherwise (in columns 1 and 2) and equals one for all turnovers and zero otherwise (in columns 3 and 4). Similarly, our specific variable of interest at this point is the interaction term between religious monitoring intensity and 2-year performance, and consistent with our hypothesis, we find a negative coefficient, which is significant at the 1% level. Again, to understand the evolution of the actual probabilities of forced turnover forced CEO turnover as the performance of the firm deteriorates, we compute the predictive margins for both the religious monitoring-intensive and the non-religious monitoring-intensive boards. Figure 3.5 graphs these margins. We find that over the 2-year holding period, both board-types do not appear to be statistically inclined to force a CEO out for matching or for barely underperforming the market up to -5%. Below this level of market-underperformance, however, we find a remarkably different type of statistically significant divergence. Specifically, while the non-religious monitoring-intensive boards remain sensitive with the probability of forced turnover rising exponentially from 0 at 0%, to close to 0.4 at -20% (with a 95% CI between close to 0.2 and close to 0.6), the probability of forced turnover rises much higher (in a concave manner) for the religious monitoring-intensive boards, from 0 at 0%, to effectively 1 at -20% (with a much tighter 95% CI).

3.5. Religiosity and CEO Compensation

In this section, we test the second of our hypotheses on the marginal returns of religiosity (or a sense of higher purpose) to intense board oversight. Specifically, we evaluate whether, compared to their non-religious counterparts, religious intense board monitors are associated with significantly lower excess CEO compensation, particularly so concerning popular pay items such as straight cash and equity grants, ceteris paribus. As in the previous section, our analyses are principally marginal to the finding that intense monitoring is associated with better board oversight (e.g., see Weisbach, 1988; Yermack, 1996; Hermalin, 2005; FHH). Indeed, apart from the hiring and firing of the CEO, the design of proper compensation contracts to adequately incentivize management is an important board monitoring function (e.g., see Browning and Sparks, 2015); and whereas, in theory, executive compensation will depend strictly on economic factors such as managerial labor market conditions and firm performance, in reality, however, management often distorts the compensation contract process for its benefits (e.g., Bechuk and Fried, 2004; FHH). Whereas FHH show that compared to their non-monitoring-intensive counterparts, monitoring-intensive boards are associated with lower excess CEO compensation, we seek to evaluate whether and how, conditional on all boards being monitoring-intensive, religiosity (or a sense of higher purpose) also matters.

Empirically, we proceed similarly to FHH. For example, we define excess compensation as the residuals from a baseline regression predicting normal compensation as a function of the economic determinants[77] of pay. We also use the natural log of total assets as a proxy for firm size and operating complexity; the ratio of the book value to the market value of equity as a proxy for growth opportunities; the market-adjusted stock return and return on asset (ROA) as proxies for firm performance; and the standard deviations of the proxies for firm performance (over the preceding five years) as proxies firm risk. We extract compensation data from the ExecuComp database; define total compensation as the natural logarithm of the sum of salary, bonus, the value of stock options and restricted stock granted during the year, long-term incentive payouts, and other miscellaneous compensation amounts; equity compensation as the natural logarithm of (one plus) the value of stock options and restricted stock awarded during the year; and cash compensation as the natural logarithm of salary plus cash bonus.

Panel A of Table 3.6 presents the results of the baseline regressions. Columns 1, 2, and 3 show the results for total, equity, and cash compensations, respectively. Consistent with previous research[78], we observe positive associations between all three compensation items and firm size, which are respectively significant at the 1% level. We similarly observe negative (or positive) associations between all three compensation items and book/market

[77] The economic determinants of pay are extracted from standard economic theory and include firm size, operating complexity, growth opportunities, firm performance, and firm risk (e.g., see also Rosen, 1982; Core, Holthausen, and Larcker, 1999).

[78] FHH, in particular

(or market/book) ratio, which are significant at the 1% level for total and equity compensations and at the 10% level for cash compensation. We also broadly observe some significant positive associations between CEO compensation items and the proxies for both firm performance and firm risk (e.g., ROA and the standard deviation of stock return).

Next, we proceed to evaluate the relationships between the modeled excess pay items (i.e., the residuals of the respective baseline regressions) and the religiosity of monitoring-intensive boards. However, while FHH used a data sample that spans the period from 1998 to 2006, our base sample runs from 1998 to 2018 (i.e., some additional 12 years; and much further from the introduction of the Sarbanes-Oxley Act, or SOX, in 2002) and over 83% of our firm-years come from the period post-2006. Consequently, we take the same approach as in the previous section and work in two steps. First, we look at monitoring-intensive boards and excess CEO pay; and second, we look at religious-monitoring-intensive boards and excess CEO pay. We discuss these steps in detail below.

3.5.1. Monitoring-intensive boards and excess CEO pay

Table A.2 (in appendix 3.IA, accompanied by granular details of the results) presents the regressions explaining excess compensation. Overall, our findings are consistent with the previous FHH results. However, it appears that that for the additional 12-year period post-2006 (i.e., for the firm-years much further from the enactment of the SOX) contained within our data sample, although monitoring intensity alone continues to, on average, significantly reduce excess equity compensation, the reductions have not been huge enough to similarly impact excess total CEO compensation, all other things equal.

3.5.2. Religious-monitoring-intensive boards and excess CEO pay

Next, we return to our main hypothesis, which is to evaluate whether the religiosity of the monitoring-intensive boards also matters for the reduction of excess CEO pay. Expectedly, we begin by using the Milbourn-Wabara measure to sort the monitoring-intensive boards into the religious and non-religious cohorts. Empirically, we simply run similar regressions to those in Panel B of Table A.3 (in Appendix 3.IA), with the only difference being that the monitoring-intensive board indicator variable is replaced with the religious monitoring-intensive board indicator variable. We show the results in Panel B of Table 3.6. Recall that the religious monitoring-intensive board variable equals one when, on a monitoring-intensive board, a majority of the straddling from a base (or initial) principal monitoring committee to another principal monitoring committee is done by the religious independent directors, and zero when, on a monitoring-intensive board, a majority of the said straddling is done by the non-religious independent directors.

We find that compared to their non-religious counterparts, religious monitoring-intensive boards are, on average, associated with a reduction in CEO cash compensation that is significant at the 1% level and a reduction in CEO equity compensation that is significant at the 5% level. However, both reductions appear to lead to a reduction in total CEO compensation that is only significant at the 10% level. While these results are generally consistent with our hypothesis, they also suggest that, compared to their non-religious counterparts, although religious intense board monitors effectively reduce both the cash and the equity components of CEO compensation, similar evidence that they tend to further reduce excess total CEO compensation is not equally as statistically significant.

3.6. Religiosity and Earnings Quality

In this section, we test the third of our hypotheses on the marginal returns of religiosity (or a sense of higher purpose) to intense board oversight. Specifically, we evaluate whether, compared to their non-religious counterparts, religious intense board monitors are more likely to, on average, further reduce earnings management, all other things equal. As in the previous two sections, our analyses are principally marginal to the finding that intense monitoring is associated with better board oversight (e.g., see Weisbach, 1988; Yermack, 1996; Hermalin, 2005; FHH). Indeed, apart from hiring and firing the CEO and/or designing proper compensation contracts to adequately incentivize management, ensuring the quality of information presented in firms' financial reports is also a vital board monitoring function (e.g., see Klein, 2002; Xie, Davidson, and DaDalt, 2003; FHH; Browning and Sparks, 2015). Whereas CKLP show that religious CEOs are associated with reduced earnings management and FHH show that, compared to non-monitoring-intensive boards, monitoring-intensive boards are also associated with reduced earnings management, we seek to evaluate whether and how, conditional on all boards being monitoring-intensive, the religiosities of monitoring-intensive boards (or their senses of higher purpose) also matter for firms' earnings quality.

Following previous research (e.g., see Jones, 1991; Dechow, Sloan, and Sweeney, 1995), we use discretionary accruals as a proxy for earnings management. Accrual accounting allows firms to measure and report their performance by recognizing economic events when they happen rather than when they make or receive the payments associated with those events. An unintended consequence, however, is that the accrual process also provides significant opportunities for firms to infuse bias into their financial statements (e.g. see FHH). Such bias often shows up in the discretionary accruals, the absolute value of abnormal accruals in particular (e.g., see Kothari, Leone, and Wasley, 2005; FHH;

CKLP). Thus, similar to FHH, we estimate two variants of the Jones (1991) model for discretionary accruals. We also control for factors—such as the external busyness of the directors, the size of the board, board independence, audit committee independence, firm size, book/market, absolute change in net income, loss, and leverage—that can affect the size of the discretionary accruals (e.g., see DeFond and Jiambalvo, 1994; Dechow, Sloan, and Sweeney, 1995; McNichols, 2000; Klein, 2002; Larcker, Richardson, and Tuna, 2007).

Again, recall that whereas FHH used a data sample that spans the period from 1998 to 2006, our base sample runs from 1998 to 2018 (i.e., some additional 12 years; and significantly after the introduction of the Sarbanes-Oxley Act, or SOX, in 2002), with over 83% of our firm-years drawn from the period post-2006. Consequently, we take the same approach as in the previous two sections—i.e., to proceed in two steps. First, we look at monitoring-intensive boards and earnings quality; and second, we look at religious-monitoring-intensive boards and earnings quality. We discuss these steps in detail below.

3.6.1. Monitoring-intensive boards and earnings quality

Table A.4 (in Appendix 3.IA, accompanied by the precise empirical specifications and granular details of the results) presents the associated regressions using our longer horizon data sample (that spans the 21 years from 1998 to 2018). In summary, even for the additional 12-year period post-2006, monitoring-intensive boards continue to be associated with economically significant reductions in earnings management.

3.6.2. Religious-monitoring-intensive boards and earnings quality

Next, we return to our hypothesis, that is, to further evaluate whether the religiosities of the intense monitors also matter for the quality of earnings information contained in firms' financial reports. Empirically, we run similar regressions to the ones in Table A.4. The only difference is that the monitoring-intensive board indicator variable is replaced with the religious monitoring-intensive board indicator variable. The religious monitoring-intensive board and other variables are as previously defined. Table 3.7 presents the results. Consistent with our hypothesis, we find that the coefficients for the religious monitoring-intensive board indicator variable are negative and statistically significant at the 5% level, across the board. This result implies that relative to their non-religious counterparts, religious intense board monitors are associated with significantly reduced earnings management. The magnitudes of the coefficients also point to some highly economically significant improvements in earnings quality. For example, the smallest coefficient in column 1 shows that as the intensity of monitoring becomes spearheaded by the religious directors, the ratio of discretionary accruals to total assets is lower by 7.61%. Accordingly, given that the mean value of discretionary accruals in the subsample is approximately 11.1% of total assets, this effectively represents an economically significant 68.6% reduction in abnormal accruals, all other things equal.

Consistent also with CKLP, we find that the religiosity of the CEO also matters for the quality of earnings information contained in firms' financial reports. Specifically, we find negative coefficients that are also statistically significant at the 5% level, across the board, for the religious CEO variable. However, the relative sizes of the coefficients suggest that, on average, the intense monitoring activities of the independent directors may matter nearly twice more, economically speaking, for the reduction of discretionary accruals, all other things equal. The results for other variables such as firm size, board size/ independence, audit committee independence, absolute change in net income all remain consistent with past research (e.g., see Klein, 2002; Larcker, Richardson, and Tuna, 2007).

3.7. Robustness Tests

In this section, we discuss the typical endogeneity concerns that can be associated with our study. We also discuss how our empirical framework and choice of testing strategies help to mitigate such problems. We conduct additional tests to rule out other plausible explanations for our results and, finally, develop and test a new hypothesis based on the Wabara (2021a) finding that implies that our results should be compatibly stronger when the same type of intense monitors are powerful or influential on the board.

3.7.1. Endogeneity concerns

The development of our novel measure (i.e., the Milbourn-Wabara measure) and our broad empirical framework derive mainly from the adaptations and extensions of the definitions and methods used in FHH and the refinement of a proxy used in CKLP. Thus, it should be unsurprising that much of the endogeneity concerns associated with our study will have been raised and discussed in the respective research. For example, it is envisagable that some measurement errors might be embedded in our proxies or measures and that possibilities for reverse causality may exist. Specifically, our proxy for the personal religiosities of the directors might be noisy (e.g., see CKLP) or the benefits of intense board oversight could have been endogenously engineered because, perhaps, the past predicaments faced by the firms/boards necessitated the intense monitoring anyway (e.g., see FHH) or that the religious directors might have been attracted to conservative firms that would typically be associated with the kinds of results we find (e.g., see CKLP).

First, regarding the possibility that some measurement errors might be embedded in our proxy for personal religiosity, it is indeed possible that, on the one hand, directors who are non-religious might have attended church-affiliated academic institutions; and on the other hand, religious directors might have attended secular academic institutions.

However, and as similarly highlighted in CKLP, to the extent that the measurement errors are not systematically and homogenously associated with the respective measures for CEO turnover, earnings management, and excess CEO compensation, the noisiness of the religiosity proxy works against the general consistency we find in our results. Moreso (and as emphasized in our rationale for refining the proxy for directors' religiosities, in section 3.2.3), attendance in church-affiliated institutions captures the pre-college religiosity and the avowed infusion of religious values into students' educational experience.

Next, regarding the reverse causality concerns, although FHH effectively address the first (i.e., that the benefits of intense board oversight could have been endogenously engineered because, perhaps, the predicaments the firms/boards faced necessitated the intense monitoring anyway) using even a quasi-experiment, we further mitigate this concern in our study by conditioning our comparative analyses on boards being already monitoring-intensive. In other words, our marginal analyses principally evaluate whether there are differences in approach and benefits that arise from the religious inclinations of the intense monitors, in particular, irrespective of whether the board's monitoring intensity is endogenously engineered or not. For the second reverse causality concern (i.e., that the religious directors might have been attracted to conservative firms that would typically be associated with the kinds of results we find), our empirical strategy also allows us to conduct additional tests to rule this out. We also conduct more tests to rule out other minor but plausible explanations for our results. We discuss these steps in detail below.

3.7.2. Additional tests

In this subsection, we present three sets of additional tests. The first set of test rules out the reverse causality concern (not ruled out in previous research) that the religious directors might have been attracted to "conservative" firms typically associated with the kinds of results we find. The second set of tests effectively rules out the possibilities that substituting the religiosity of the intense board monitors with that of any of the chairmen of the principal monitoring committees, that of a majority of the chairmen of the principal monitoring committees, or that of the lead independent director could reproduce our results, across the board. The third develops a new hypothesis to test the application of the Wabara (2021a) finding that suggests that our results should be compatibly stronger when the same type of intense monitors are powerful or influential on the board

3.7.2.1. Religiosity of monitoring-intensive directors or "conservativeness" of the firms?

Testing our main hypotheses up to this point, we have systematically used the generalized Milbourn-Wabara measure conditioned only on boards being monitoring-intensive. However, as schematized in Figure 3.6, if the second reverse causality concern (i.e., that the religious intense board monitors might have been attracted to conservative firms typically associated with the kinds of results we find) were true, then we should also find similar results using a version of the generalized Milbourn-Wabara measure conditioned only on boards being non-monitoring-intensive. Mathematically, if our results were driven by the conservativeness of the firms, then it should not matter whether we use:

$$MWm_{i,t} = \begin{cases} 0, & MWm_{i,t,0} = \mu_{i,t}^{R_g} < 0,\ FHHm_{i,t} = 1 \\ 1, & MWm_{i,t,0} = \mu_{i,t}^{R_g} > 0,\ FHHm_{i,t} = 1 \end{cases}$$

Or we use:

$$MWm_{i,t}{}' = \begin{cases} 0, & MWm_{i,t,0} = \mu_{i,t}^{R_g} < 0,\ FHHm_{i,t} = 0 \\ 1, & MWm_{i,t,0} = \mu_{i,t}^{R_g} > 0,\ FHHm_{i,t} = 0 \end{cases}$$

Where $MWm_{i,t}{}'$ is a similar duo-conditional rendering of our measure that states that a corporate board for firm i is a non-religious non-monitoring-intensive or a non-religious non-intense board monitor in year t (i.e., $MWm_{i,t}{}' = 0$) if the board is non-monitoring-intensive (i.e., $FHHm_{i,t} = 0$) and a higher number of the straddling across the three principal monitoring committees is spearheaded by the non-religious independent directors (i.e., $MWm_{i,t,0} = \mu_{i,t}^{R_g} < 0$). Conversely, it also states that a corporate board for firm i is religious non-monitoring-intensive or a religious non-intense board monitor in year t (i.e., $MWm_{i,t}{}' = 1$) if the board is non-monitoring-intensive (i.e., $FHHm_{i,t} = 0$) and a higher number of the straddling across the three principal monitoring committees is spearheaded by the religious independent directors (i.e., $MWm_{i,t,0} = \mu_{i,t}^{R_g} > 0$).

The logic behind this empirical test is straightforward. If our religion-intense monitoring measures mainly capture the endogenous conservativeness of the firms, then merely substituting $MWm_{i,t}$ with $MWm_{i,t}{}'$ in our empirical models should produce similar results (at least, in terms of direction and statistical significance). Otherwise, our results must be due to the religiosity of the intense monitors, and we would have effectively ruled out the second reverse causality concern. Tables 3.15 and 3.16 show the key regression results. The coefficients for the religious non-monitoring-intensive board variable ($MWm_{i,t}{}'$) are all not only directionally inconsistent but also statistically insignificant, across the board, even at the 10% level. Consequently, we rule out this reverse causality concern.

3.7.2.2. Monitoring intensity or board/committee position?

Having ruled out the second reverse causality concern (that the religious directors might be attracted to conservative firms typically associated with the kinds of results we find), we return to our initial parametrization of the Wabara-Wilbourn measures for this study (i.e., $MWm_{i,t}$) to, in this case, further test whether substituting the religiosity of the intense board monitors with that of any of the chairmen of the principal monitoring committees, that of a majority of the chairmen of the principal monitoring committees, or that of the lead independent director, could effectively reproduce our results.

Tables 3.8 to 3.13 present the regression results. Specifically, Table 3.8 shows that our results on the sensitivity of CEO turnover to 1-year firm performance are due primarily to the religiosity of the intense board monitors. Table 3.9 shows that our results on the sensitivity of CEO turnover to 2-year firm performance are due primarily to the religiosity of the intense board monitors. Similarly, Tables 3.10 to 3.12 show that our results on the reduction of excess CEO compensation are due primarily to the religiosity of the intense board monitors. Table 3.13 also shows that our results on the reduction of earnings management are due primarily to the religiosity of the intense board monitors. We nonetheless find weak hints that the effects of the religious intense board monitors might be boosted when the lead independent director and/or a majority of the principal monitoring committee chairs are also religious. The latter is consistent with the findings of Wabara (2021a). Hence we develop and test the additional hypothesis discussed below.

3.7.3. Religiosity backed by power, and influence

Wabara (2021a) suggests that "within-group or within-board power and influence distributions impact whether individual tendencies become manifest in small group (or board) settings." This finding implies that if indeed our results are driven by the religiosity of the monitoring-intensive directors, then our results should be stronger when the same type of intense monitors are also powerful or influential on the board. Recall from Section 3.5 that the evidence that compared to their non-religious counterparts, religious monitoring-intensive directors tend to further reduce excess total CEO compensation was not equally as statistically strong (compared to other related results). Hence, we hypothesize as follows.

Hypothesis 4: *Compared to a non-religious monitoring-intensive board with a non-religious lead independent director and/or on which a majority of the chairmen of the principal monitoring committees are non-religious, a religious monitoring-intensive board with a religious lead independent director and/or on which a majority of the chairmen of the principal monitoring committees are religious will, on average, significantly reduce excess total CEO compensation, all other things equal.*

To empirically evaluate this potential amplification of the marginal effects of religiosity when the religiosity of the monitoring-intensive directors is backed by some compatible power and/or influence, we create a new variable called powerful religious monitoring-intensive board. We do this by further conditioning the Milbourn-Wabara measure ($MWm_{i,t}$) on both the religiosities of the lead independent director and those of a majority of the chairmen of the principal monitoring committees. Mathematically, we set:

$$\widetilde{MWm}_{i,t} = \begin{cases} 0, & MWm_{i,t} = 0,\ C_1 = 0,\ C_2 = 0 \\ 1, & MWm_{i,t} = 1,\ C_1 = 1,\ C_2 = 1 \end{cases}$$

Where $\widetilde{MWm}_{i,t}$ is a tri-conditional rendering of our measure that states that a corporate board for firm i is powerful non-religious monitoring-intensive in year t (i.e., $\widetilde{MWm}_{i,t} = 0$) if the board is non-religious monitoring-intensive (i.e., $MWm_{i,t} = 0$) and both the lead independent director and a majority of the chairmen of the principal monitoring committees are also non-religious (i.e., $C_1 = 0$ and $C_2 = 0$, respectively). Conversely, it also states that a corporate board for firm i is powerful religious monitoring-intensive in year t (i.e., $\widetilde{MWm}_{i,t} = 1$) if the board is religious monitoring-intensive (i.e., $MWm_{i,t} = 1$) and both the lead independent director and a majority of the chairmen of the principal monitoring committees are also religious (i.e., $C_1 = 1$ and $C_2 = 1$, respectively).

In other words, to test this hypothesis, we simply replace the religious monitoring-intensive indicator variable ($MWm_{i,t}$) with the powerful religious monitoring-intensive indicator variable ($\widetilde{MWm}_{i,t}$), while keeping all else unchanged in all of our regression models. For precision, powerful religious monitoring-intensive board equals one when the board is religious monitoring-intensive, the lead independent director is religious, and a majority of the chairmen of the principal monitoring committees are religious, zero when the board is non-religious monitoring-intensive, the lead independent director is non-religious, and a majority of the chairmen of the principal monitoring committees are non-religious. The regression results provide strong support for this hypothesis. For example, Table 3.14 shows that consistent with this additional hypothesis, the evidence that, relative to their non-religious counterparts, religious intense board monitors tend to further reduce excess total CEO compensation, previously not equally statistically strong, becomes highly significant when the lead independent director and/or a majority of the principal monitoring committee chairs are also religious. When extended to other aspects of our findings, this hypothesis's main idea again proves true (not tabulated).

3.8. Conclusion

We begin by outlining the parallels and potential linkages between personal religiosity and a higher sense of purpose. Next, we assume that a proxy for personal religiosity exists and then adapt/extend the FHH definitions for intense monitoring to develop a novel two-dimensional measure to capture monitoring-intensive directors' aggregate religious inclinations. We generalize our novel measure and, for ease of reference, call it the Milbourn-Wabara measure. However, to activate this measure as a practical sorting device, we adapt a good proxy for personal religiosity. CKLP first used a version of this proxy. Next, we use the activated Milbourn-Wabara measure to examine whether there exist any marginal returns of religiosity (or a sense of higher purpose) to intense board oversight. Specifically, we evaluate whether, relative to their non-religious counterparts, religious monitoring-intensive directors exhibit, on average, greater sensitivity of CEO turnover to firm performance, lower propensity to overcompensate the CEO, and/or lower likelihood of being associated with poor earnings quality.

We find that, compared to their non-religious counterparts, religious intense board monitors exhibit significantly lower sensitivity of CEO turnover to firm performance over a holding period of 1 year. However, for the more extended holding period of 2 years, this difference in sensitivity significantly switches direction, consistent with the "*higher purpose, incentives, and economic performance"* theory, which suggests that believers in higher purpose will tend to hold a longer-term perspective. We also find that religious monitoring-intensive directors further significantly reduce both earnings management and excess total CEO compensation, especially when the lead independent director and/or a majority of the principal monitoring committee chairs are also religious.

We also affirm, consistent with previous research, that CEOs' religiosities matter for the quality of earnings information in firms' financial reports. However, we further show that the intense board monitors' religiosity also matters, perhaps even more so, economically speaking. Finally, we rule out the potential reverse causality concern, previously not ruled out by past research, that religious directors might be systematically attracted to "conservative" firms that would typically be associated with the kind of results we find. Overall, our findings show that religious monitoring-intensive directors differentially influence intense board oversight results and, thereby, help infuse or propagate a corporate culture consistent with an authentic organizational higher purpose.

In terms of opportunities for future work, our generalized Milbourn-Wabara measure can be a useful device to continue to examine the impacts of religiosity (or a sense of higher purpose) on the effectiveness of strategic board advising; firm values; firm bankruptcies; et cetera. Moreover, the main ideas behind our measure's construction can also be adapted and extended to new contexts to study multiple other phenomena (in finance, corporate governance in particular). We, for example, plan to continue to develop and explore new research ideas along these lines.

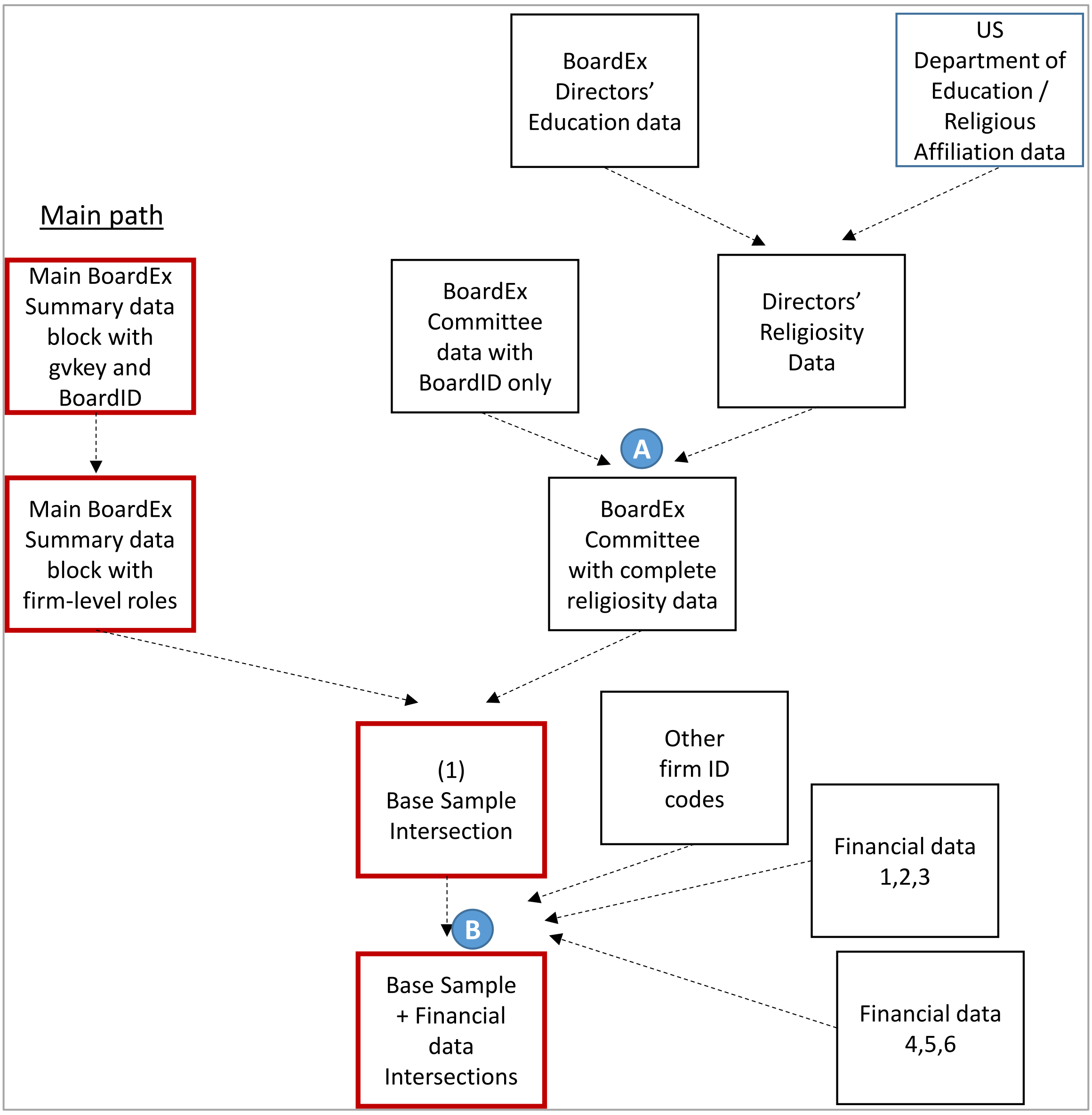


Figure 3.1
Generating our base sample from the raw data
We highlight the main data path on the left using red boxes: we begin with the BoardEx summary data block with both firm and board identification numbers and generate variables that capture firm-level roles, et cetera. However, we start our data intersections from the top right. First, we intersect the BoardEx directors' education data with the schools/religious affiliation data from the U.S. Department of Education to generate a data block that assigns religiosity to all directors based on our proxy (as described in section 3.2.3). Next, we intersect the directors' religiosity data block with the BoardEx committee data block (with only board identification numbers) to generate the BoardEx committee data. The resultant data block has complete religiosity information for every board committee member. In other words, we keep 100% of the firm-years for which we can assign religiosity to every individual on the board and call this point A. We randomly corroborate our biographical data by looking up directors' information on LinkedIn, Bloomberg, company, and other biographical websites. Next, we intersect the committee data (with complete religiosity information for every board member) with the summary data containing firm-level variables to generate our base sample (pre-intersection with the various accounting and financial data) and call this point B. Next, we intersect the latter with the respective financial and accounting data to generate our base sample post-intersection with financial data. Finally, we evaluate whether any firm-years for which we could not assign religiosity to every member of the board (at point A) could have made it into our final data (at point B). Interestingly, we find that none of such firm-years could have, and thus have no impact on our empirical analyses.

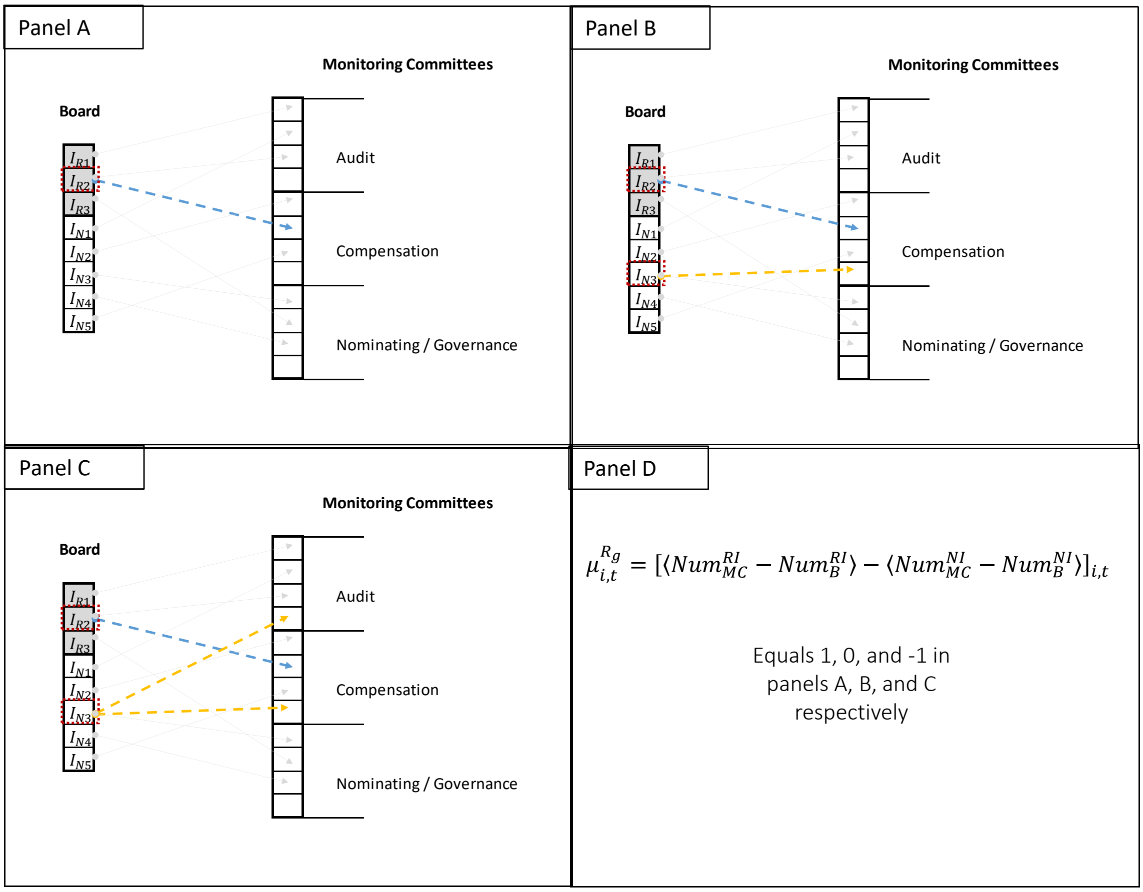


Figure 3.2
Computation/generalization of the religion-intense monitoring parameter, $\mu_{i,t}^{R_g}$

This figure schematizes the basic methodology for computing the religion-intense monitoring parameter ($\mu_{i,t}^{R_g}$). Let panels A, B, and C represent three different years for the board of firm i; and I_{jk} the k^{th} independent director of religious-type j. Hence, in each year, there are 8 independent directors on the board, 3 of whom are religious (R) and 5 of whom are non-religious (N). First, each independent director is assigned a base (or initial) monitoring committee (i.e., the gray arrows). In each of the panels, there is at least one independent director that serves in more than one monitoring committee. Such independent directors are highlighted using a red dashed box and are monitoring-intensive (according to the FHH definition). In panel A, there is only one monitoring-intensive independent director on the board and she is religious. She straddles between her base (Audit) and the Compensation committees. Consequently, the total number of monitoring committee straddles from the base committees is 1 for the religious independent directors (i.e., $\langle Num_{MC}^{RI} - Num_{B}^{RI}\rangle = 1$) and 0 for the non-religious independent directors (i.e., $\langle Num_{MC}^{NI} - Num_{B}^{NI}\rangle = 0$). Thus $\mu_{i,t}^{R_g} = 1$ and we say that the intense monitoring for this panel or year is religious. Panel D shows that, following a similar logic, we compute $\mu_{i,t}^{R_g} = 0$ and $\mu_{i,t}^{R_g} = -1$ for panels B and C respectively; and we say that the intense board monitorings for those panels or years are religion-indifferent and non-religious, respectively. Nevertheless, to allow for the possibility of a coarse aggregation of our firm-period religion-intense monitoring parameters, we define a generalized version as follows:

$$MWm_{i,t,n} = \frac{\sum_{\tau=0}^{n} \omega_{t-\tau}\mu_{i,t-\tau}^{R_g}}{n+1}$$

Where, for ease of reference, we call $MWm_{i,t,n}$ the Milbourn-Wabara measure for the religious inclination of the intense board monitors for firm i, in period t, aggregated over n preceding periods up to period t. n is an integer that ideally may be less than 2 but could also depend on some desired technical meaning, the level of coarseness sought, or the minimum number of firm-periods per firm contained within the entire sample. $\omega_{t-\tau}$ is a weighting factor that may depend on the importance of the $(t-\tau)th$ period to the period-t event. If $n = 0$, and $\omega_t = 1$ for all t, then $MWm_{i,t,n} = MWm_{i,t,0} = \mu_{i,t}^{R_g}$ and is strictly equivalent to a firm-period sample.

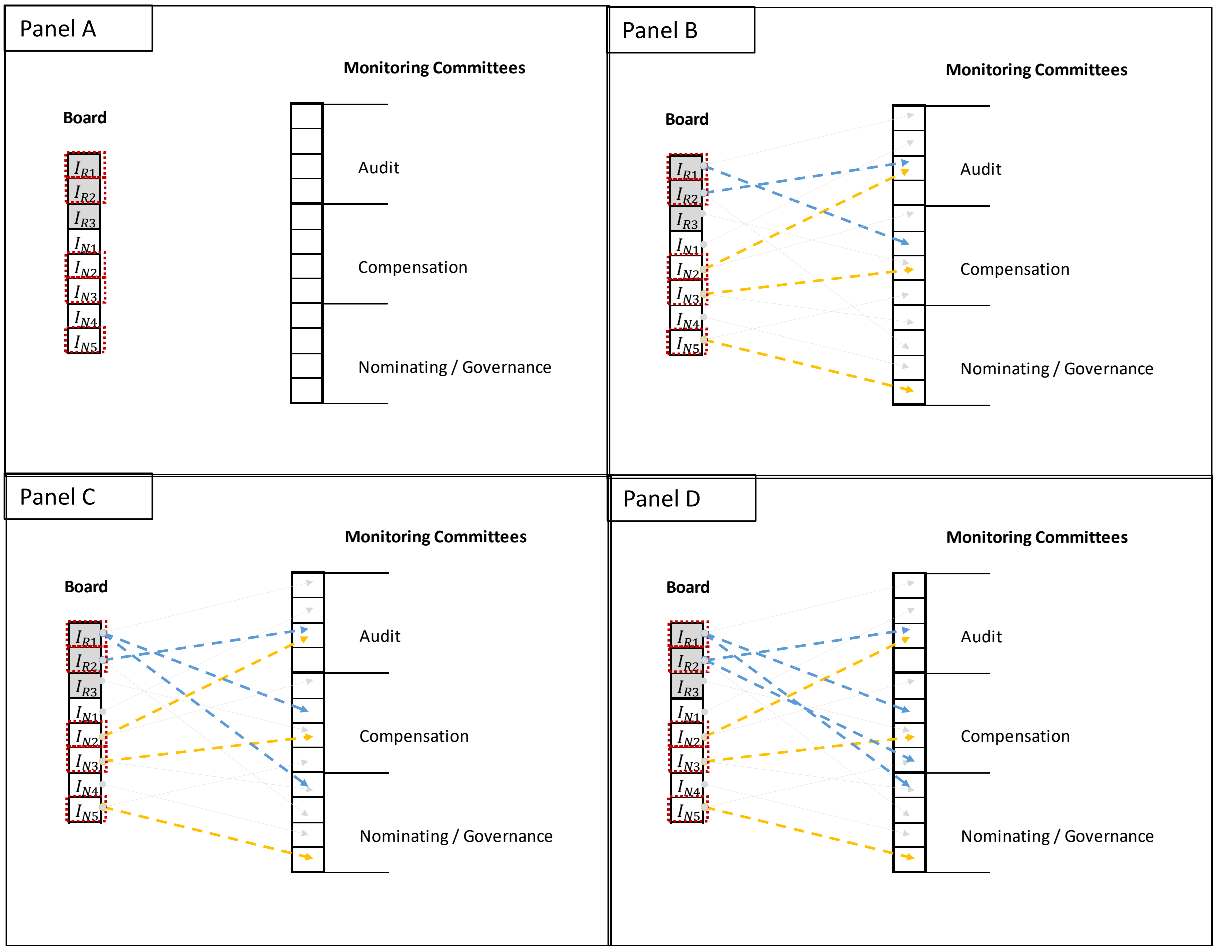


Figure 3.3
Construction of the Milbourn-Wabara measure ($MWm_{i,t}$) for this study

This figure schematizes our construction of the generalized Milbourn-Wabara measure ($MWm_{i,t,n}$) for the religious inclination of the intense board monitors for firm i, in period t, aggregated over n preceding periods up to period t, where t is in years (e.g., see Figure 3.2). In this study, we take a simplified cum marginal approach relative to FHH. Specifically, we set $n = 0$ and $\omega_t = 1$ for all t, such that:

$$MWm_{i,t} = \begin{cases} 0, & MWm_{i,t,0} = \mu_{i,t}^{R_g} < 0,\ FHHm_{i,t} = 1 \\ 1, & MWm_{i,t,0} = \mu_{i,t}^{R_g} > 0,\ FHHm_{i,t} = 1 \end{cases}$$

For example, let panels B, C, and D represent three different years for the board of firm i; and I_{jk} the k^{th} independent director of religious-type j. Hence, in each year, there are 8 independent directors on the board, 3 of whom are religious (R) and 5 of whom are non-religious (N). Panel A accentuates that the condition $FHHm_{i,t} = 1$ is always satisfied for all the panels since a majority (i.e., 5/8) of the independent directors are monitoring-intensive. These directors are highlighted using a red dashed box. This also means that the entire board is monitoring-intensive in each of the panels or years. Each director is, first, assigned a base (or initial) monitoring committee (i.e., the gray arrows) and the monitoring-intensive directors straddle between their base and some other committees (i.e., the dashed blue and orange arrows, for the religious and non-religious independent directors, respectively). In panel B, $MWm_{i,t,0} = \mu_{i,t}^{R_g} < 0$, hence $MWm_{i,t} = 0$, and we call this a non-religious monitoring-intensive board; the directors, non-religious intense board monitors; and their monitoring, non-religious intense board monitoring. Conversely, In panel D, $MWm_{i,t,0} = \mu_{i,t}^{R_g} > 0$, hence $MWm_{i,t} = 1$, and we call this a religious monitoring-intensive board; the directors, religious intense board monitors; and their monitoring, religious intense board monitoring. In panel C, however, $MWm_{i,t,0} = \mu_{i,t}^{R_g} = 0$, hence $MWm_{i,t}$ is undefined for our comparative purposes and is unused.

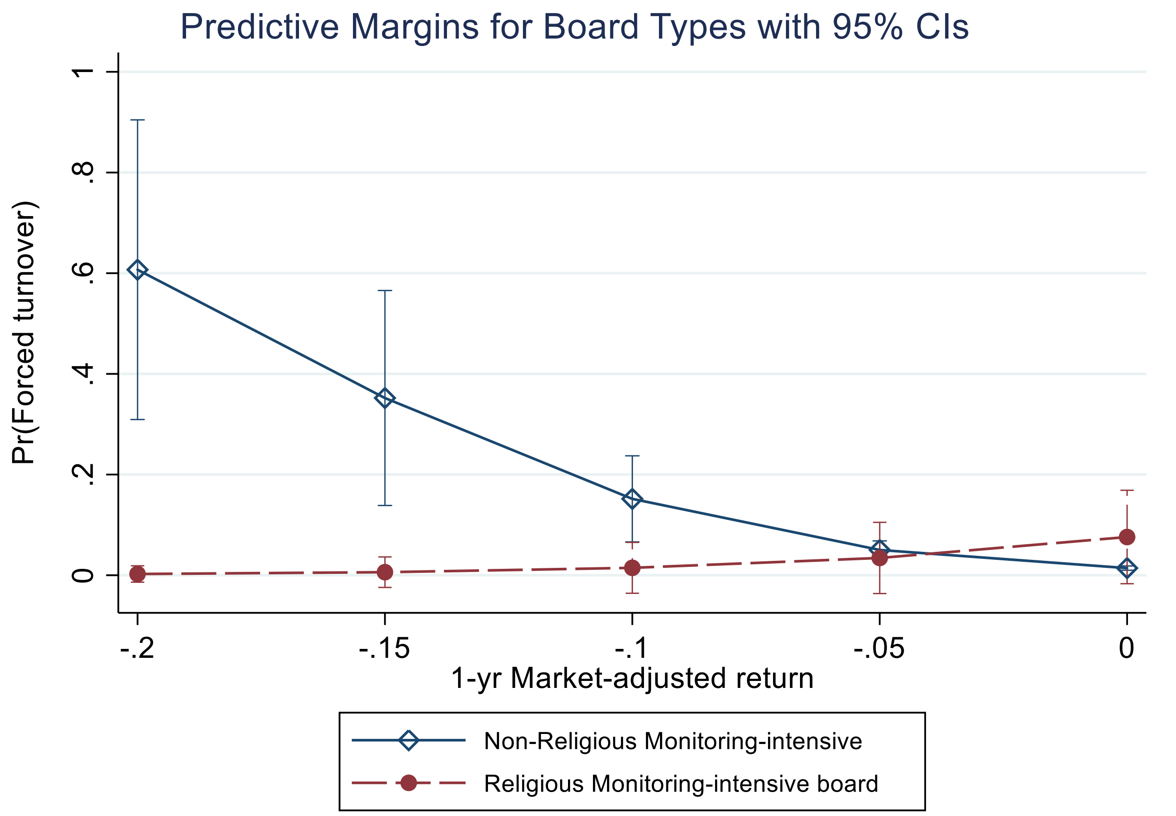


Figure 3.4
Predictive margins of religious monitoring-intensive boards and 1-yr performance
Predictive margins are statistics calculated from predictions of a previously fit model at fixed values of some covariates and averaging or otherwise integrating over the remaining covariates (e.g., see Buis, 2010; Hosmer, Lemeshow, and Sturdivant, 2013; Long and Freese, 2014; Norton, Wang, and Ai, 2004; Williams, 2012). The previously fit model is the logistic regression model in Table 3.4, column 1. The vertical axis shows the actual probabilities of forced turnover. The horizontal axis shows the 1-year firm performance or market-adjusted return (defined as the annual stock return less same period return on the CRSP value-weighted portfolio of NYSE/Amex/Nasdaq stocks). The continuous navy blue line with a hollow diamond represents the average sensitivity of forced CEO turnover to 1-year firm performance for the non-religious monitoring-intensive boards. The dashed maroon line with solid maroon circles represents the average sensitivity of forced CEO turnover to 1-year firm performance for religious monitoring-intensive boards. Both lines have vertical outlines of the 95% CI.

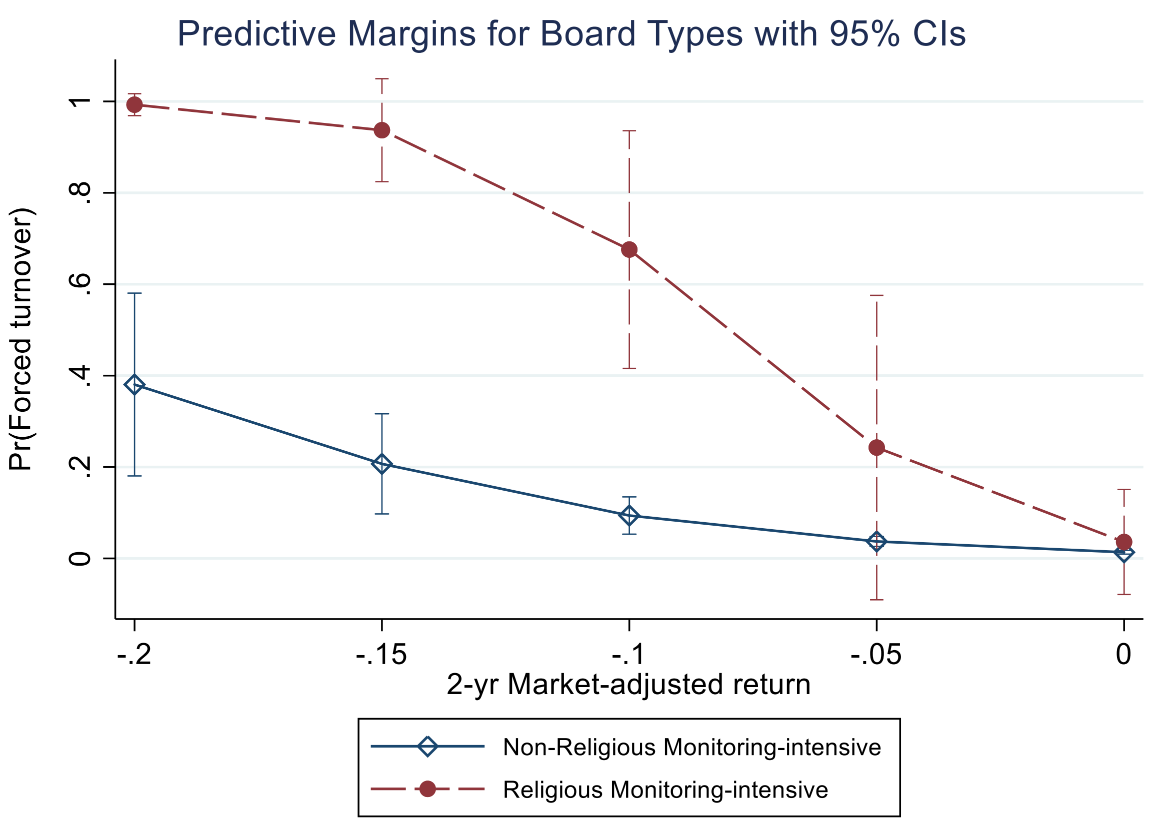


Figure 3.5
Predictive margins of religious monitoring-intensive boards and 2-yr performance
Predictive margins are statistics calculated from predictions of a previously fit model at fixed values of some covariates and averaging or otherwise integrating over the remaining covariates (e.g., see Buis, 2010; Hosmer, Lemeshow, and Sturdivant, 2013; Long and Freese, 2014; Norton, Wang, and Ai, 2004; Williams, 2012). The previously fit model is the logistic regression model in Table 3.5, column 1. The vertical axis shows the actual probabilities of forced turnover. The horizontal axis shows the 2-year firm performance or market-adjusted return (defined as the bi-annual stock return less same period return on the CRSP value-weighted portfolio of NYSE/Amex/Nasdaq stocks). The continuous navy blue line with a hollow diamond represents the average sensitivity of forced CEO turnover to 2-year firm performance for the non-religious monitoring-intensive boards. The dashed maroon line with solid maroon circles represents the average sensitivity of forced CEO turnover to 2-year firm performance for religious monitoring-intensive boards. Both lines have vertical outlines of the 95% CI.

Religious Monitoring-intensive board variable

	Religious	Non-Religious
Monitoring-Intensive	$MWm_{i,t,0} = \mu_{i,t}^{R_g} > 0$ $FHHm_{i,t} = 1$	$MWm_{i,t,0} = \mu_{i,t}^{R_g} < 0$ $FHHm_{i,t} = 1$
Non-Monitoring-Intensive	$MWm_{i,t,0} = \mu_{i,t}^{R_g} > 0$ $FHHm_{i,t} = 0$	$MWm_{i,t,0} = \mu_{i,t}^{R_g} < 0$ $FHHm_{i,t} = 0$

$$MWm_{i,t} = \begin{cases} 0, & MWm_{i,t,0} = \mu_{i,t}^{R_g} < 0,\ FHHm_{i,t} = 1 \\ 1, & MWm_{i,t,0} = \mu_{i,t}^{R_g} > 0,\ FHHm_{i,t} = 1 \end{cases}$$

$$MWm_{i,t}{}' = \begin{cases} 0, & MWm_{i,t,0} = \mu_{i,t}^{R_g} < 0,\ FHHm_{i,t} = 0 \\ 1, & MWm_{i,t,0} = \mu_{i,t}^{R_g} > 0,\ FHHm_{i,t} = 0 \end{cases}$$

Religious Non-Monitoring-intensive board variable

Figure 3.6
Addressing a reverse causality concern

A potential reverse causality concern for this study is that the religious intense board monitors might have been attracted to conservative firms typically associated with the kinds of results we find. However, if the religious and non-religious directors are simply attracted to conservative and non-conservative firms, respectively, then replacing the Religious Monitoring-intensive board variable ($MWm_{i,t}$) with the Religious Non-Monitoring-intensive board variable ($MWm_{i,t}{}'$) in our analyses should not matter for our results. The logic is straightforward: If our religion-intense monitoring parameters mainly capture the endogenous conservativeness of the firms, then merely substituting $MWm_{i,t}$ with $MWm_{i,t}{}'$ in our empirical models should produce similar results (at least, in terms of direction and statistical significance). Otherwise, our results must be due to the religiosity of the intense monitors, and we would have effectively ruled out this reverse causality concern.

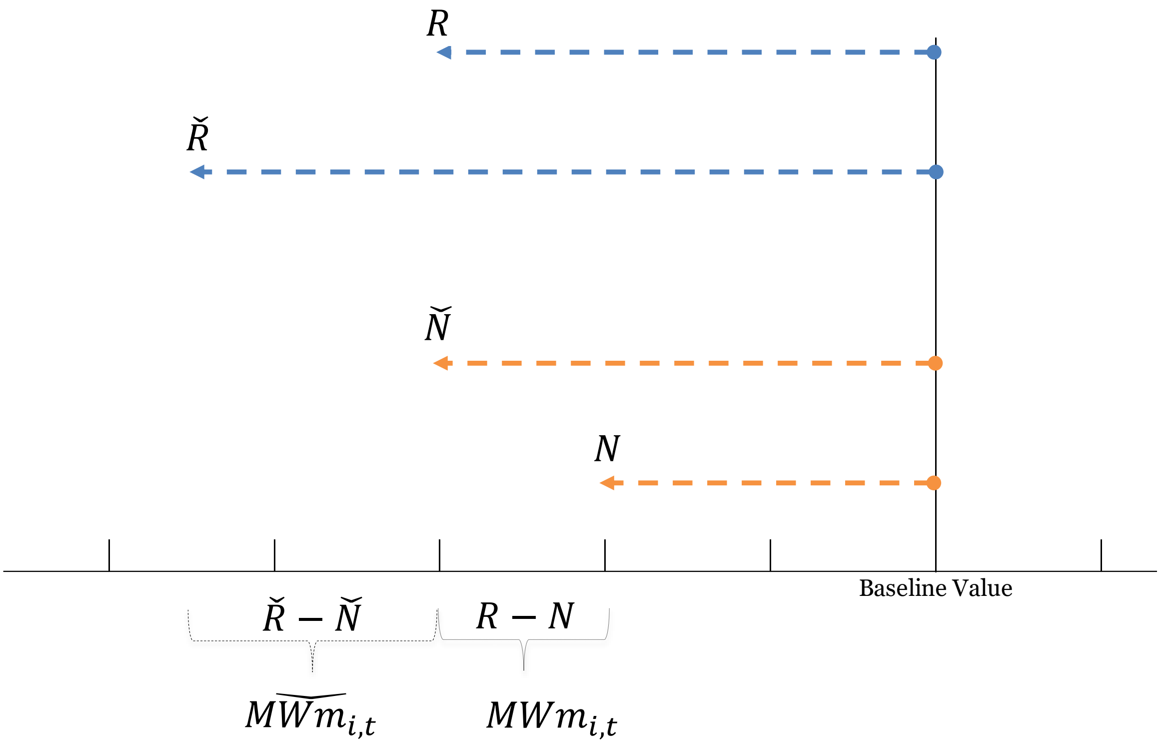


Figure 3.7
Construction of the Powerful Milbourn-Wabara measure ($\widetilde{MWm}_{i,t}$) for this study
Let N and R be the reductive effects of the non-religious and religious directors, respectively, from a baseline value when they are not backed by the power and/or influence of the lead independent director and a majority of the chairs of the principal monitoring committee with similar religiosity on the board. Then $MWm_{i,t}$ measures the difference $R - N$ for a given firm in a given year. Now, suppose that backing either group by power and influence amplifies the reductive effects by 150% to $\breve{N}$ and $\breve{R}$, respectively, then $\widetilde{MWm}_{i,t}$ measures the difference $\breve{R} - \breve{N}$. Thus, we can create a Powerful religious monitoring-intensive board variable as follows:

$$\widetilde{MWm}_{i,t} = \begin{cases} 0, & MWm_{i,t} = 0,\ C_1 = 0,\ C_2 = 0 \\ 1, & MWm_{i,t} = 1,\ C_1 = 1,\ C_2 = 1 \end{cases}$$

Which states that a corporate board for a firm i is:

- Powerful non-religious monitoring-intensive in year t (i.e., $\widetilde{MWm}_{i,t} = 0$) if the board is non-religious monitoring-intensive (i.e., $MWm_{i,t} = 0$) and both the lead independent director and a majority of the chairmen of the principal monitoring committees are also non-religious (i.e., $C_1 = 0$ and $C_2 = 0$, respectively)
- Powerful religious monitoring-intensive in year t (i.e., $\widetilde{MWm}_{i,t} = 1$) if the board is religious monitoring-intensive (i.e., $MWm_{i,t} = 1$) and both the lead independent director and a majority of the chairmen of the principal monitoring committees are also religious (i.e., $C_1 = 1$ and $C_2 = 1$, respectively)

.

Wabara (2021) suggests that while $MWm_{i,t}$ and $\widetilde{MWm}_{i,t}$ will lead to similar results (i.e., directionally), those from the latter will be stronger or, at least, more statistically significant.

Table 3.1
Frequency distribution of the BoardEx sectors in the base sample
The base data consist of a total of 37 sectors, of which none overly dominates the entire sample. Of the 23,528 firm-years, only three sectors come the closest to 10% and the rest have between 0.4% and 6.9%, each.

Fim-year distribution by BoardEx sectors

Sector	Frequency	Percent	Cummulative (%)
Aerospace & Defence	267	1.1	1
Automobiles & Parts	343	1.5	3
Beverages	95	0.4	3
Business Services	1,062	4.5	8
Chemicals	613	2.6	10
Clothing & Personal Products	326	1.4	12
Construction & Building Materials	584	2.5	14
Consumer Services	172	0.7	15
Containers & Packaging	113	0.5	15
Diversified Industrials	172	0.7	16
Education	103	0.4	16
Electricity	174	0.7	17
Electronic & Electrical Equipment	2,152	9.2	26
Engineering & Machinery	1,014	4.3	31
Food & Drug Retailers	85	0.4	31
Food Producers & Processors	469	2.0	33
Forestry & Paper	183	0.8	34
General Retailers	704	3.0	37
Health	1,616	6.9	44
Household Products	353	1.5	45
Information Technology Hardware	590	2.5	48
Leisure & Hotels	687	2.9	50
Leisure Goods	90	0.4	51
Media & Entertainment	330	1.4	52
Mining	385	1.6	54
Oil & Gas	1,413	6.0	60
Pharmaceuticals and Biotechnology	2,870	12.2	72
Publishing	166	0.7	73
Real Estate	1,534	6.5	79
Renewable Energy	231	1.0	80
Software & Computer Services	1,903	8.1	88
Speciality & Other Finance	1,119	4.8	93
Steel & Other Metals	225	1.0	94
Telecommunication Services	607	2.6	97
Tobacco	6	0.0	97
Transport	562	2.4	99
Wholesale Trade	210	0.9	100
Total	23,528	100	

Table 3.2
Annual distribution of the principal monitoring committees and the Milbourn-Wabara Measures

The base sample consists of 23,528 firm-years of board characteristics, for 2,945 unique firms, from 1998 to 2018. Audit refers to the percentage of boards in the year with an audit committee. A similar definition applies to Compensation, and Nominating/Governance committees. Monitoring-intensive refers to the percentage of firms in the year with a monitoring-intensive board. Non-religious refers to the percentage of firms in the year with non-religious boards, based on the definition of the Milbourn-Wabara measures. A similar definition to the latter applies to Non-religious Religion-indifferent, and Religious.

Year	Sample	Audit	Compensation	Nominating/Governance	Monitoring-intensive	Non-religious	Religion-indifferent	Religious
1998	1	100%	100%	100%	0%	100%	0%	0%
1999	35	100%	97%	49%	23%	80%	11%	9%
2000	283	100%	94%	42%	34%	81%	15%	4%
2001	316	100%	94%	52%	35%	84%	11%	5%
2002	356	100%	95%	68%	47%	88%	8%	4%
2003	607	100%	95%	75%	55%	91%	6%	3%
2004	711	100%	95%	88%	59%	92%	5%	3%
2005	788	100%	95%	89%	61%	92%	5%	3%
2006	892	100%	95%	88%	58%	92%	6%	2%
2007	1,099	99%	93%	83%	55%	88%	10%	2%
2008	1,169	99%	93%	83%	57%	87%	11%	2%
2009	1,208	99%	93%	85%	59%	88%	10%	2%
2010	1,269	100%	94%	87%	59%	90%	9%	1%
2011	1,347	99%	94%	87%	59%	90%	9%	1%
2012	1,480	99%	94%	86%	58%	90%	9%	1%
2013	1,642	100%	94%	87%	58%	90%	9%	2%
2014	1,796	100%	95%	88%	58%	90%	8%	2%
2015	1,964	100%	95%	88%	59%	90%	8%	2%
2016	2,101	100%	95%	89%	58%	91%	7%	2%
2017	2,205	100%	96%	91%	60%	93%	6%	1%
2018	2,259	100%	97%	93%	59%	93%	5%	1%
All years	23,528	100%	95%	86%	58%	90%	8%	2%

Table 3.3
Board, firm, and CEO characteristics.

Summary statistics for all firm-years after intersecting the base sample with the financial and accounting data from Compustat, Execucomp, etc.

Variable	Mean	Median	25th Percentile	75th Percentile	Standard deviation
Board characteristics					
Directors	7	6	5	8	3
Independent directors	6	6	4	8	3
Board independence	0.9	1.0	1.0	1.0	0.3
Externally busy board	0.0	0.0	0.0	0.0	0.1
Audit committee membership	4	4	3	4	1
Compensation committee membership	4	3	3	4	2
Nominating/governance committee membership	4	3	3	5	2
Audit committee independence	1.0	1.0	1.0	1.0	0.2
Compensation committee independence	0.9	1.0	1.0	1.0	0.3
Nominating/governance committee independence	0.8	1.0	1.0	1.0	0.4
Firm characteristics					
Total assets	5395	697	124	3300	15492
Return on assets	0.0	0.1	0.0	0.2	0.5
Market value of equity	4401	739	146	3064	9893
Book/Market	0.4	0.4	0.2	0.7	0.6
CEO characteristics					
CEO age	55.7	56.0	50.0	61.0	8.1
CEO duality	0.4	0.0	0.0	1.0	0.5
CEO directors	0.2	0.2	0.0	0.4	0.2
CEO ownership **	2.2	0.4	0.1	1.4	5.1
Total compensation	8.2	8.3	7.6	9.0	1.0

*** (in %)*

Table 3.4

Religiosity of monitoring intensity and sensitivity of CEO turnover to 1-year market-adjusted firm performance.

The dependent variable in columns 1 and 2 equals one for firm-years with forced CEO turnovers, zero otherwise. The dependent variable in columns 3 and 4 equals one for firm-years with any CEO turnovers, zero otherwise. Religious monitoring-intensive board equals one when, on a monitoring-intensive board, a majority of the straddling from a base (or initial) principal monitoring committee to another principal monitoring committee is done by the religious independent directors, zero when, on a monitoring-intensive board, a majority of the said straddling is done by the non-religious independent directors. 1-year market-adjusted return is annual stock return less same period return on the CRSP value-weighted portfolio of NYSE/Amex/Nasdaq stocks. CEO duality equals one when the CEO also serves as board chair, zero otherwise. Externally busy board equals one when a majority of independent directors serve on three or more boards. Board size is the natural log of the number of directors. Institutional ownership is the percentage of outstanding shares owned by institutional investors. CEO ownership is the proportion of outstanding shares beneficially owned by the CEO. Firm size is the natural log of total assets. CEO age is measured in years. Board independence equals one when a majority of directors are independent, zero otherwise. Religious CEO equals one if the CEO is religious, zero otherwise. Each regression includes year and sector fixed effects. Numbers in parentheses are p-values based on robust standard errors clustered at the firm level. Levels of significance are indicated by ***, **, and * for 1%, 5%, and 10%, respectively.

	Forced Turnover		*All Turnover*	
	(1)	*(2)*	*(3)*	*(4)*
Religious monitoring-intensive board	1.8609 (0.7796)**	1.8490 (0.7686)**	0.9766 (0.5110)*	0.9774 (0.5112)*
1-yr Market-adjusted return	-27.3985 (4.5817)***	-27.6890 (4.6735)***	-9.0732 (2.1845)***	-9.0644 (2.1837)***
Religious monitoring-intensive board x 1-yr Market-adjusted return	45.3236 (16.5002)***	45.9844 (16.5079)***	10.0386 (11.2867)	10.0398 (11.2901)
CEO duality	-0.9985 (0.3602)***	-0.9654 (0.3625)***	-0.4216 (0.1413)***	-0.4215 (0.1412)***
Externally busy board	5.7757 (1.7376)***	5.7105 (1.7434)***	0.8565 (1.3924)	0.8586 (1.3920)
Board size	0.8019 (0.7698)	0.7212 (0.7812)	0.3864 (0.3086)	0.3913 (0.3098)
Institutional ownership	-1.3787 (0.8185)*	-1.3395 (0.8089)*	-0.3145 (0.3576)	-0.3157 (0.3572)
CEO ownership	-0.0652 (0.0482)	-0.0667 (0.0489)	-0.0530 (0.0218)**	-0.0529 (0.0218)**
Firm size	0.1198 (0.1163)	0.1273 (0.1169)	0.0219 (0.0442)	0.0214 (0.0443)
CEO age	0.0289 (0.0231)	0.0279 (0.0232)	0.0617 (0.0093)***	0.0617 (0.0093)***
Board independence			1.6475 (0.8380)**	1.6448 (0.8378)**
Religious CEO				0.0717 (0.3556)
Year fixed effects	Yes	Yes	Yes	Yes
Sector fixed effects	Yes	Yes	Yes	Yes
Pseudo R-Squared	0.171	0.172	0.065	0.065
N	3,244	3,193	3,918	3,918

* $p<0.1$; ** $p<0.05$; *** $p<0.01$
Standard errors in parentheses

Table 3.5

Religiosity of monitoring intensity and sensitivity of CEO turnover to 2-year market-adjusted firm performance.

The dependent variable in columns 1 and 2 equals one for firm-years with forced CEO turnovers, zero otherwise. The dependent variable in columns 3 and 4 equals one for firm-years with any CEO turnovers, zero otherwise. Religious monitoring-intensive board equals one when, on a monitoring-intensive board, a majority of the straddling from a base (or initial) principal monitoring committee to another principal monitoring committee is done by the religious independent directors, zero when, on a monitoring-intensive board, a majority of the said straddling is done by the non-religious independent directors. 2-year market-adjusted return is bi-annual stock return less same period return on the CRSP value-weighted portfolio of NYSE/Amex/Nasdaq stocks. CEO duality equals one when the CEO also serves as board chair, zero otherwise. Externally busy board equals one when a majority of independent directors serve on three or more boards. Board size is the natural log of the number of directors. Institutional ownership is the percentage of outstanding shares owned by institutional investors. CEO ownership is the proportion of outstanding shares beneficially owned by the CEO. Firm size is the natural log of total assets. CEO age is measured in years. Board independence equals one when a majority of directors are independent, zero otherwise. Religious CEO equals one if the CEO is religious, zero otherwise. Each regression includes year and sector fixed effects. Numbers in parentheses are p-values based on robust standard errors clustered at the firm level. Levels of significance are indicated by ***, **, and * for 1%, 5%, and 10%, respectively.

	Forced Turnover		*All Turnover*	
	(1)	*(2)*	*(3)*	*(4)*
Religious monitoring-intensive board	1.1259 (1.8308)	1.0985 (1.8454)	1.0313 (0.5286)*	1.0316 (0.5287)*
2-yr Market-adjusted return	-21.8845 (3.2626)***	-22.2200 (3.3628)***	-7.4202 (1.6908)***	-7.4132 (1.6947)***
Religious monitoring-intensive board x 2-yr Market-adjusted return	-31.2922 (14.9888)**	-30.7697 (14.9871)**	-6.1819 (7.8083)	-6.1897 (7.8111)
CEO duality	-1.0570 (0.3607)***	-1.0338 (0.3654)***	-0.4452 (0.1443)***	-0.4450 (0.1442)***
Externally busy board	5.6724 (1.7827)***	5.5879 (1.7727)***	0.8581 (1.3948)	0.8591 (1.3946)
Board size	0.5976 (0.7619)	0.5004 (0.7752)	0.3676 (0.3125)	0.3703 (0.3141)
Institutional ownership	-0.9159 (0.8178)	-0.8522 (0.8164)	-0.2155 (0.3781)	-0.2165 (0.3780)
CEO ownership	-0.0701 (0.0551)	-0.0711 (0.0560)	-0.0510 (0.0222)**	-0.0510 (0.0223)**
Firm size	0.1473 (0.1250)	0.1590 (0.1269)	0.0292 (0.0452)	0.0289 (0.0455)
CEO age	0.0355 (0.0227)	0.0348 (0.0229)	0.0636 (0.0095)***	0.0636 (0.0095)***
Board independence			2.4250 (1.0866)**	2.4230 (1.0869)**
Religious CEO				0.0377 (0.3878)
Year fixed effects	Yes	Yes	Yes	Yes
Sector fixed effects	Yes	Yes	Yes	Yes
Pseudo R-Squared	0.202	0.203	0.071	0.071
N	3,164	3,115	3,819	3,819

* $p<0.1$; ** $p<0.05$; *** $p<0.01$
Standard errors in parentheses

Table 3.6

Religious intense monitoring and excess CEO compensation

Panel A is the same as in Table A.3 (of Appendix 3.IA) and presents regressions predicting normal CEO compensation as a function of the economic determinants of executive pay during 1998–2018. Total compensation is the natural log of the sum of salary, bonus, the value of stock options and restricted stock granted during the year, long-term incentive payouts, and other miscellaneous compensation amounts. Equity compensation is the natural log of (one plus) the value of stock options and restricted stock awarded during the year. Cash compensation is the natural log of salary plus cash bonus. Firm size is the natural log of total assets. Book/market is the book value of equity divided by the market value of equity. Stock return is the annual stock return less same-period return on the CRSP value-weighted portfolio of NYSE/Amex/Nasdaq stocks. ROA is operating income before depreciation divided by total assets. SDRet and SDROA are the respective standard deviations of stock return and ROA over the preceding five years. Each regression includes year and sector fixed effects.

Panel B presents regressions explaining excess compensation, defined as the residuals from the respective Panel A regressions. The dependent variables in all columns are the actual residuals from the respective Panel A regressions. Here, however, Religious monitoring-intensive board equals one when, on a monitoring-intensive board, a majority of the straddling from a base (or initial) principal monitoring committee to another principal monitoring committee is done by the religious independent directors, zero when, on a monitoring-intensive board, a majority of the said straddling is done by the non-religious independent directors. Board size is the natural log of the number of directors. CEO duality equals one when the CEO also serves as board chair, zero otherwise. CEO directors is the number of directors who are CEOs of other firms. Religious CEO equals one if the CEO is religious, zero otherwise. Numbers in parentheses are p-values based on robust standard errors clustered at the firm level. Levels of significance are indicated by ***, **, and * for 1%, 5%, and 10%, respectively.

Panel A: Economic determinants

	Total Compensation	*Equity Compensation*	*Cash Compensation*
Firm size	0.4057 (0.0176)***	0.5054 (0.0157)***	0.1607 (0.0148)***
Book/market	-0.1735 (0.0612)***	-0.1938 (0.0717)***	-0.0572 (0.0330)*
Stock return	0.2307 (0.4093)	-0.3681 (0.4158)	-0.2230 (0.3520)
ROA	0.4180 (0.1578)***	0.5158 (0.2261)**	0.1921 (0.1495)
SDRet	0.2764 (0.3130)	0.9382 (0.4382)**	-0.3363 (0.2461)
SDROA	0.8105 (1.0557)	2.1660 (1.3146)*	-1.8941 (1.0803)*
Year fixed effects	Yes	Yes	Yes
Sector fixed effects	Yes	Yes	Yes
R-Squared	0.517	0.565	0.259
N	8,058	6,632	8,072

Panel B: Excess compensation

	Total Compensation		*Equity Compensation*		*Cash Compensation*	
	(1)	*(2)*	*(1)*	*(2)*	*(1)*	*(2)*
Religious Monitoring-intensive board	-0.3794 (0.1997)*	-0.3773 (0.1999)*	-0.4927 (0.2310)**	-0.4892 (0.2309)**	-0.2311 (0.0710)***	-0.2303 (0.0712)***
Board size	0.2429 (0.0842)***	0.2458 (0.0844)***	0.1785 (0.1245)	0.1832 (0.1240)	0.1297 (0.0940)	0.1307 (0.0943)
CEO duality	0.0839 (0.0403)**	0.0839 (0.0402)**	-0.0253 (0.0560)	-0.0252 (0.0560)	0.0681 (0.0344)**	0.0681 (0.0344)**
CEO directors	-0.0013 (0.0916)	-0.0002 (0.0916)	0.0446 (0.1356)	0.0464 (0.1358)	-0.0237 (0.0871)	-0.0234 (0.0872)
Religious CEO		0.0773 (0.0924)		0.0988 (0.1642)		0.0275 (0.0629)
R-Squared	0.015	0.016	0.006	0.006	0.008	0.009
N	4,141	4,141	3,385	3,385	4,145	4,145

* $p<0.1$; ** $p<0.05$; *** $p<0.01$
Standard errors in parentheses

Table 3.7

Religious intense monitoring and earnings quality

The dependent variable in the first and second columns is the absolute value of discretionary accruals generated from the modified Jones model over 1998 – 2018. The dependent variable in the third and fourth columns is the absolute value of discretionary accruals generated from the modified Jones model augmented with a control for firm performance as in Kothari, Leone, and Wasley (2005). Religious monitoring-intensive board equals one when, on a monitoring-intensive board, a majority of the straddling from a base (or initial) principal monitoring committee to another principal monitoring committee is done by the religious independent directors, zero when, on a monitoring-intensive board, a majority of the said straddling is done by the non-religious independent directors. Board independence is the percentage of directors that are unaffiliated with the firm beyond their directorship. Audit committee independence is the percentage of independent directors on the audit committee. Firm size is the natural log of market value of equity. Book/market is the book value of equity divided by the market value of equity. The absolute change in net income is the absolute value of the change in net income between years $t - 1$ and t. Leverage is the ratio of total assets to liabilities. Numbers in parentheses are p-values based on robust standard errors clustered at the firm level. Levels of significance are indicated by ***, **, and * for 1%, 5%, and 10%, respectively.

	Abnormal accruals		*ROA augmented abnormal accruals*	
	(1)	*(2)*	*(3)*	*(4)*
Religious Monitoring-intensive board	-0.0761 (0.0295)**	-0.0762 (0.0296)**	-0.0766 (0.0298)**	-0.0767 (0.0299)**
Externally busy board	0.0455 (0.0448)	0.0447 (0.0447)	0.0444 (0.0431)	0.0437 (0.0431)
Board size	-0.1273 (0.0366)***	-0.1287 (0.0368)***	-0.1277 (0.0375)***	-0.1291 (0.0377)***
Board independence	-0.1314 (0.0738)*	-0.1304 (0.0737)*	-0.1320 (0.0741)*	-0.1311 (0.0741)*
Audit committee independence	0.1011 (0.1406)	0.0992 (0.1406)	0.1018 (0.1427)	0.0998 (0.1427)
Firm size	-0.0211 (0.0058)***	-0.0211 (0.0058)***	-0.0212 (0.0059)***	-0.0212 (0.0059)***
Book/market	-0.0008 (0.0004)*	-0.0008 (0.0004)*	-0.0008 (0.0004)*	-0.0008 (0.0004)*
Absolute change in net income	0.0000 (0.0000)***	0.0000 (0.0000)***	0.0000 (0.0000)***	0.0000 (0.0000)***
Leverage	-0.0000 (0.0000)*	-0.0000 (0.0000)*	-0.0000 (0.0000)*	-0.0000 (0.0000)*
Religious CEO		-0.0485 (0.0204)**		-0.0495 (0.0208)**
R-Squared	0.042	0.042	0.041	0.041
N	4,072	4,072	4,070	4,070

* p<0.1; ** p<0.05; *** p<0.01
Standard errors in parentheses

Table 3.8

Robustness tests: Religious intense monitoring, CEO turnover, and 1-yr Market-adjusted return

The dependent variable in all columns equals one for firm-years with forced CEO turnovers, zero otherwise. Religious monitoring-intensive board equals one when, on a monitoring-intensive board, a majority of the straddling from a base (or initial) principal monitoring committee to another principal monitoring committee is done by the religious independent directors, zero when, on a monitoring-intensive board, a majority of the said straddling is done by the non-religious independent directors. Religious lead independent director equals one when the lead independent director is religious, zero otherwise. Religious audit chair equals one when the chairman of the audit committee is religious, zero otherwise. Religious compensation chair equals one when the chairman of the compensation committee is religious, zero otherwise. Religious nominating and governance chair equals one when the chairman of the nominating and governance committee is religious, zero otherwise. Religious majority monitoring chair equals one when a majority of the chairmen of the principal monitoring committees are religious, zero otherwise. 1-year market-adjusted return is annual stock return less same period return on the CRSP value-weighted portfolio of NYSE/Amex/Nasdaq stocks. CEO duality equals one when the CEO also serves as board chair, zero otherwise. Institutional ownership is the percentage of outstanding shares owned by institutional investors. CEO ownership is the proportion of outstanding shares beneficially owned by the CEO. Firm size is the natural log of total assets. CEO age is measured in years. Each regression includes year and sector fixed effects. Numbers in parentheses are p-values based on robust standard errors clustered at the firm level. Levels of significance are indicated by ***, **, and * for 1%, 5%, and 10%, respectively.

	Forced Turnover					
	(1)	(2)	(3)	(4)	(5)	(6)
Religious monitoring-intensive board	1.8490 (0.7686)**					
Religious lead independent director		0.1088 (0.5526)				
Religious audit chair			0.4468 (0.4851)			
Religious compensation chair				0.6944 (0.4016)*		
Religious nominating and governance chair					0.0810 (0.5136)	
Religious majority monitoring chair						0.9626 (0.5531)*
1-yr Market-adjusted return	-27.6890 (4.6735)***	-25.5192 (5.0913)***	-26.8579 (4.8602)***	-26.3779 (5.2505)***	-26.5317 (4.9736)***	-26.8672 (4.6395)***
Religious monitoring-intensive board x 1-yr Market-adjusted return	45.9844 (16.5079)***					
Religious lead independent director x 1-yr Market-adjusted return		-3.4505 (12.3343)				
Religious audit chair x 1-yr Market-adjusted return			11.1859 (15.3084)			
Religious compensation chair x 1-yr Market-adjusted return				6.0838 (10.5048)		
Religious nominating and governance chair x 1-yr Market-adjusted return					6.8739 (11.7169)	
Religious majority monitoring chair x 1-yr Market-adjusted return						21.1716 (17.2834)
CEO duality	-0.9654 (0.3625)***	-0.9908 (0.3522)***	-0.9696 (0.3580)***	-0.9658 (0.3583)***	-0.9968 (0.3575)***	-0.9526 (0.3605)***
Externally busy board	5.7105 (1.7434)***	5.6518 (1.7687)***	5.6299 (1.7642)***	5.7167 (1.7702)***	5.6317 (1.7596)***	5.6337 (1.7600)***
Board size	0.7212 (0.7812)	0.5282 (0.7785)	0.5774 (0.7919)	0.4967 (0.7759)	0.5161 (0.7742)	0.5692 (0.7622)
Institutional ownership	-1.3395 (0.8089)*	-1.4682 (0.7839)*	-1.4518 (0.7881)*	-1.4408 (0.7972)*	-1.4341 (0.7807)*	-1.4766 (0.7889)*
CEO ownership	-0.0667 (0.0488)	-0.0656 (0.0462)	-0.0649 (0.0480)	-0.0632 (0.0475)	-0.0662 (0.0463)	-0.0659 (0.0481)
Firm size	0.1273 (0.1169)	0.1185 (0.1162)	0.1124 (0.1165)	0.1310 (0.1190)	0.1216 (0.1137)	0.1203 (0.1166)
CEO age	0.0279 (0.0232)	0.0305 (0.0219)	0.0309 (0.0227)	0.0280 (0.0220)	0.0311 (0.0224)	0.0311 (0.0224)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Sector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R-Squared	0.172	0.160	0.162	0.164	0.161	0.164
N	3,193	3,193	3,193	3,193	3,193	3,193

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Standard errors in parentheses

Table 3.9

Robustness tests: Religious intense monitoring, CEO turnover, and 2-yr Market-adjusted return

The dependent variable in all columns equals one for firm-years with forced CEO turnovers, zero otherwise. Religious monitoring-intensive board equals one when, on a monitoring-intensive board, a majority of the straddling from a base (or initial) principal monitoring committee to another principal monitoring committee is done by the religious independent directors, zero when, on a monitoring-intensive board, a majority of the said straddling is done by the non-religious independent directors. Religious lead independent director equals one when the lead independent director is religious, zero otherwise. Religious audit chair equals one when the chairman of the audit committee is religious, zero otherwise. Religious compensation chair equals one when the chairman of the compensation committee is religious, zero otherwise. Religious nominating and governance chair equals one when the chairman of the nominating and governance committee is religious, zero otherwise. Religious majority monitoring chair equals one when a majority of the chairmen of the principal monitoring committees are religious, zero otherwise. 2-year market-adjusted return is bi-annual stock return less same period return on the CRSP value-weighted portfolio of NYSE/Amex/Nasdaq stocks. CEO duality equals one when the CEO also serves as board chair, zero otherwise. Institutional ownership is the percentage of outstanding shares owned by institutional investors. CEO ownership is the proportion of outstanding shares beneficially owned by the CEO. Firm size is the natural log of total assets. CEO age is measured in years. Each regression includes year and sector fixed effects. Numbers in parentheses are *p*-values based on robust standard errors clustered at the firm level. Levels of significance are indicated by ***, **, and * for 1%, 5%, and 10%, respectively.

	Forced Turnover					
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Religious monitoring-intensive board	1.0985					
	(1.8454)					
Religious lead independent director		0.2311				
		(0.5810)				
Religious audit chair			0.1299			
			(0.5437)			
Religious compensation chair				0.2854		
				(0.4815)		
Religious nominating and governance chair					0.0563	
					(0.5418)	
Religious majority monitoring chair						0.0546
						(0.8688)
2-yr Market-adjusted return	-22.2200	-22.8563	-22.2435	-21.3375	-22.7300	-21.8725
	(3.3628)***	(3.5314)***	(3.4943)***	(3.6871)***	(3.3637)***	(3.3826)***
Religious monitoring-intensive board x 2-yr Market-adjusted return	-30.7697					
	(14.8871)**					
Religious lead independent director x 2-yr Market-adjusted return		-1.9191				
		(12.9676)				
Religious audit chair x 2-yr Market-adjusted return			-6.8360			
			(8.4696)			
Religious compensation chair x 2-yr Market-adjusted return				-7.8958		
				(7.7552)		
Religious nominating and governance chair x 2-yr Market-adjusted return					-3.2515	
					(8.9496)	
Religious majority monitoring chair x 2-yr Market-adjusted return						-18.7166
						(11.1624)*
CEO duality	-1.0338	-1.0093	-1.0134	-1.0437	-1.0197	-1.0161
	(0.3854)***	(0.3581)***	(0.3564)***	(0.3684)***	(0.3611)***	(0.3613)***
Externally busy board	5.5879	5.5797	5.5684	5.6250	5.5785	5.6154
	(1.7727)***	(1.7581)***	(1.7845)***	(1.7664)***	(1.7605)***	(1.7838)***
Board size	0.5004	0.3436	0.3508	0.4290	0.3240	0.4745
	(0.7752)	(0.7419)	(0.7305)	(0.7692)	(0.7432)	(0.7586)
Institutional ownership	-0.8522	-0.9510	-0.9421	-0.7774	-0.9827	-0.9119
	(0.8184)	(0.7989)	(0.7982)	(0.8324)	(0.7961)	(0.7969)
CEO ownership	-0.0711	-0.0661	-0.0660	-0.0613	-0.0674	-0.0667
	(0.0580)	(0.0488)	(0.0496)	(0.0499)	(0.0494)	(0.0510)
Firm size	0.1590	0.1532	0.1636	0.1577	0.1593	0.1407
	(0.1269)	(0.1275)	(0.1231)	(0.1269)	(0.1225)	(0.1251)
CEO age	0.0348	0.0315	0.0340	0.0289	0.0322	0.0337
	(0.0228)	(0.0219)	(0.0226)	(0.0227)	(0.0224)	(0.0225)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Sector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R-Squared	0.203	0.195	0.196	0.199	0.195	0.200
N	3,115	3,115	3,115	3,115	3,115	3,115

* $p<0.1$; ** $p<0.05$; *** $p<0.01$
Standard errors in parentheses

Table 3.10

Robustness tests: Religious intense monitoring and excess CEO compensation (cash)

The dependent variables in all columns are the actual residuals from the corresponding Table 3.8, Panel A regressions. Religious monitoring-intensive board equals one when, on a monitoring-intensive board, a majority of the straddling from a base (or initial) principal monitoring committee to another principal monitoring committee is done by the religious independent directors, zero when, on a monitoring-intensive board, a majority of the said straddling is done by the non-religious independent directors. Religious lead independent director equals one when the lead independent director is religious, zero otherwise. Religious audit chair equals one when the chairman of the audit committee is religious, zero otherwise. Religious compensation chair equals one when the chairman of the compensation committee is religious, zero otherwise. Religious nominating and governance chair equals one when the chairman of the nominating and governance committee is religious, zero otherwise. Religious majority monitoring chair equals one when a majority of the chairmen of the principal monitoring committees are religious, zero otherwise. CEO duality equals one when the CEO also serves as board chair, zero otherwise. CEO directors is the number of directors who are CEOs of other firms. Numbers in parentheses are *p*-values based on robust standard errors clustered at the firm level. Levels of significance are indicated by ***, **, and * for 1%, 5%, and 10%, respectively.

	Excess Cash Compensation					
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Religious monitoring-intensive board	-0.2311 (0.0710)***					
Religious lead independent director		-0.0083 (0.0567)				
Religious audit chair			-0.0441 (0.0643)			
Religious compensation chair				-0.0041 (0.0570)		
Religious nominating and governance chair					-0.0254 (0.0420)	
Religious majority monitoring chair						-0.0941 (0.0738)
Board size	0.1297 (0.0940)	0.1375 (0.0935)	0.1359 (0.0929)	0.1373 (0.0936)	0.1408 (0.0948)	0.1358 (0.0929)
CEO duality	0.0681 (0.0344)**	0.0696 (0.0345)**	0.0691 (0.0345)**	0.0698 (0.0345)**	0.0696 (0.0345)**	0.0685 (0.0346)**
CEO directors	-0.0237 (0.0871)	-0.0271 (0.0875)	-0.0247 (0.0886)	-0.0273 (0.0872)	-0.0265 (0.0876)	-0.0243 (0.0878)
R-Squared	0.008	0.007	0.008	0.007	0.007	0.008
N	4,145	4,145	4,145	4,145	4,145	4,145

* $p<0.1$; ** $p<0.05$; *** $p<0.01$
Standard errors in parentheses

Table 3.11

Robustness tests: Religious intense monitoring and excess CEO compensation (equity)

The dependent variables in all columns are the actual residuals from the corresponding Table 3.8, Panel A regressions. Religious monitoring-intensive board equals one when, on a monitoring-intensive board, a majority of the straddling from a base (or initial) principal monitoring committee to another principal monitoring committee is done by the religious independent directors, zero when, on a monitoring-intensive board, a majority of the said straddling is done by the non-religious independent directors. Religious lead independent director equals one when the lead independent director is religious, zero otherwise. Religious audit chair equals one when the chairman of the audit committee is religious, zero otherwise. Religious compensation chair equals one when the chairman of the compensation committee is religious, zero otherwise. Religious nominating and governance chair equals one when the chairman of the nominating and governance committee is religious, zero otherwise. Religious majority monitoring chair equals one when a majority of the chairmen of the principal monitoring committees are religious, zero otherwise. CEO duality equals one when the CEO also serves as board chair, zero otherwise. CEO directors is the number of directors who are CEOs of other firms. Numbers in parentheses are p-values based on robust standard errors clustered at the firm level. Levels of significance are indicated by ***, **, and * for 1%, 5%, and 10%, respectively.

	Excess Equity Compensation					
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Religious monitoring-intensive board	-0.4927 (0.2310)**					
Religious lead independent director		-0.0588 (0.0686)				
Religious audit chair			0.0160 (0.0897)			
Religious compensation chair				0.0046 (0.0765)		
Religious nominating and governance chair					0.0654 (0.0592)	
Religious majority monitoring chair						-0.0310 (0.0906)
Board size	0.1785 (0.1245)	0.1982 (0.1234)	0.1964 (0.1236)	0.1962 (0.1235)	0.1861 (0.1237)	0.1961 (0.1235)
CEO duality	-0.0253 (0.0560)	-0.0247 (0.0560)	-0.0237 (0.0560)	-0.0238 (0.0561)	-0.0234 (0.0559)	-0.0241 (0.0561)
CEO directors	0.0446 (0.1356)	0.0437 (0.1359)	0.0393 (0.1355)	0.0399 (0.1357)	0.0375 (0.1354)	0.0407 (0.1356)
R-Squared	0.006	0.004	0.004	0.004	0.004	0.004
N	3,385	3,385	3,385	3,385	3,385	3,385

* $p<0.1$, ** $p<0.05$, *** $p<0.01$
Standard errors in parentheses

Table 3.12

Robustness tests: Religious intense monitoring and excess CEO compensation (total)

The dependent variables in all columns are the actual residuals from the corresponding Table 3.8, Panel A regressions. Religious monitoring-intensive board equals one when, on a monitoring-intensive board, a majority of the straddling from a base (or initial) principal monitoring committee to another principal monitoring committee is done by the religious independent directors, zero when, on a monitoring-intensive board, a majority of the said straddling is done by the non-religious independent directors. Religious lead independent director equals one when the lead independent director is religious, zero otherwise. Religious audit chair equals one when the chairman of the audit committee is religious, zero otherwise. Religious compensation chair equals one when the chairman of the compensation committee is religious, zero otherwise. Religious nominating and governance chair equals one when the chairman of the nominating and governance committee is religious, zero otherwise. Religious majority monitoring chair equals one when a majority of the chairmen of the principal monitoring committees are religious, zero otherwise. CEO duality equals one when the CEO also serves as board chair, zero otherwise. CEO directors is the number of directors who are CEOs of other firms. Numbers in parentheses are p-values based on robust standard errors clustered at the firm level. Levels of significance are indicated by ***, **, and * for 1%, 5%, and 10%, respectively.

	Excess Total Compensation					
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Religious monitoring-intensive board	-0.3794 (0.1997)*					
Religious lead independent director		-0.0435 (0.0523)				
Religious audit chair			0.0064 (0.0553)			
Religious compensation chair				-0.0219 (0.0659)		
Religious nominating and governance chair					-0.0318 (0.0501)	
Religious majority monitoring chair						-0.1313 (0.0722)*
Board size	0.2429 (0.0842)***	0.2564 (0.0832)***	0.2558 (0.0833)***	0.2553 (0.0835)***	0.2599 (0.0843)***	0.2534 (0.0828)***
CEO duality	0.0839 (0.0403)**	0.0858 (0.0404)**	0.0868 (0.0404)**	0.0866 (0.0404)**	0.0864 (0.0404)**	0.0848 (0.0404)**
CEO directors	-0.0013 (0.0916)	-0.0055 (0.0916)	-0.0079 (0.0914)	-0.0065 (0.0899)	-0.0063 (0.0918)	-0.0030 (0.0913)
R-Squared	0.015	0.013	0.013	0.013	0.013	0.014
N	4,141	4,141	4,141	4,141	4,141	4,141

* $p<0.1$; ** $p<0.05$; *** $p<0.01$
Standard errors in parentheses

Table 3.13

Robustness tests: Religious intense monitoring and earnings quality

The dependent variable in all columns is the absolute value of discretionary accruals generated from the modified Jones model over 1998 – 2018. Religious monitoring-intensive board equals one when, on a monitoring-intensive board, a majority of the straddling from a base (or initial) principal monitoring committee to another principal monitoring committee is done by the religious independent directors, zero when, on a monitoring-intensive board, a majority of the said straddling is done by the non-religious independent directors. Religious lead independent director equals one when the lead independent director is religious, zero otherwise. Religious audit chair equals one when the chairman of the audit committee is religious, zero otherwise. Religious compensation chair equals one when the chairman of the compensation committee is religious, zero otherwise. Religious nominating and governance chair equals one when the chairman of the nominating and governance committee is religious, zero otherwise. Religious majority monitoring chair equals one when a majority of the chairmen of the principal monitoring committees are religious, zero otherwise. Externally busy board equals one when a majority of independent directors serve on three or more boards. Board size is the natural log of the number of directors. Board independence is the percentage of directors that are unaffiliated with the firm beyond their directorship. Audit committee independence is the percentage of independent directors on the audit committee. Firm size is the natural log of market value of equity. Book/market is the book value of equity divided by the market value of equity. The absolute change in net income is the absolute value of the change in net income between years $t-1$ and t. Leverage is the ratio of total assets to liabilities. Numbers in parentheses are p-values based on robust standard errors clustered at the firm level. Levels of significance are indicated by ***, **, and * for 1%, 5%, and 10%, respectively.

	Abnormal Accruals					
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Religious monitoring-intensive board	-0.0762					
	(0.0296)**					
Religious lead independent director		-0.0407				
		(0.0117)***				
Religious audit chair			-0.0192			
			(0.0120)			
Religious compensation chair				-0.0014		
				(0.0260)		
Religious nominating and governance chair					0.0020	
					(0.0234)	
Religious majority monitoring chair						0.0306
						(0.0582)
Externally busy board	0.0447	0.0430	0.0437	0.0453	0.0457	0.0466
	(0.0447)	(0.0448)	(0.0449)	(0.0449)	(0.0450)	(0.0449)
Board size	-0.1287	-0.1279	-0.1268	-0.1271	-0.1271	-0.1268
	(0.0368)***	(0.0367)***	(0.0366)***	(0.0366)***	(0.0366)***	(0.0367)***
Board independence	-0.1304	-0.1284	-0.1292	-0.1299	-0.1299	-0.1297
	(0.0737)*	(0.0736)*	(0.0736)*	(0.0736)*	(0.0737)*	(0.0738)*
Audit committee independence	0.0992	0.1009	0.0961	0.0975	0.0972	0.0955
	(0.1406)	(0.1406)	(0.1403)	(0.1408)	(0.1407)	(0.1406)
Firm size	-0.0211	-0.0208	-0.0210	-0.0210	-0.0211	-0.0212
	(0.0058)***	(0.0057)***	(0.0058)***	(0.0058)***	(0.0059)***	(0.0058)***
Book/market	-0.0008	-0.0008	-0.0008	-0.0008	-0.0008	-0.0008
	(0.0004)*	(0.0004)*	(0.0004)*	(0.0004)*	(0.0004)*	(0.0004)*
Absolute change in net income	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)***
Leverage	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
	(0.0000)*	(0.0000)*	(0.0000)*	(0.0000)*	(0.0000)*	(0.0000)*
Religious CEO	-0.0485	-0.0476	-0.0445	-0.0484	-0.0484	-0.0479
	(0.0204)**	(0.0202)**	(0.0201)**	(0.0207)**	(0.0204)**	(0.0203)**
R-Squared	0.042	0.042	0.041	0.041	0.041	0.041
N	4,072	4,072	4,072	4,072	4,072	4,072

* $p<0.1$; ** $p<0.05$; *** $p<0.01$
Standard errors in parentheses

Table 3.14

Robustness tests: Powerful religious intense monitoring and excess CEO compensation

The dependent variables in all columns are the actual residuals from the corresponding Table 3.8, Panel A regressions. Powerful religious monitoring-intensive board equals one when the board is religious monitoring-intensive, the lead independent director is religious, and a majority of the chairmen of the principal monitoring committees are religious, zero when the board is non-religious monitoring-intensive, the lead independent director is non-religious, and a majority of the chairmen of the principal monitoring committees are non-religious. CEO duality equals one when the CEO also serves as board chair, zero otherwise. CEO directors is the number of directors who are CEOs of other firms. Numbers in parentheses are p-values based on robust standard errors clustered at the firm level. Levels of significance are indicated by ***, **, and * for 1%, 5%, and 10%, respectively.

Excess CEO compensation	*Total Compensation*		*Equity Compensation*		*Cash Compensation*	
	(1)	*(2)*	*(1)*	*(2)*	*(1)*	*(2)*
Powerful religious monitoring-intensive board	-0.2474 (0.0639)***	-0.2443 (0.0642)***	-0.4751 (0.0963)***	-0.4697 (0.0957)***	-0.3699 (0.0708)***	-0.3691 (0.0711)***
Board size	0.2370 (0.0857)***	0.2401 (0.0859)***	0.2046 (0.1304)	0.2104 (0.1297)	0.1115 (0.0941)	0.1122 (0.0943)
CEO duality	0.0872 (0.0432)**	0.0872 (0.0431)**	-0.0240 (0.0601)	-0.0239 (0.0600)	0.0677 (0.0372)*	0.0677 (0.0372)*
CEO directors	0.0096 (0.0965)	0.0107 (0.0964)	0.1010 (0.1429)	0.1031 (0.1431)	-0.0264 (0.0910)	-0.0262 (0.0910)
Religious CEO		0.0772 (0.0980)		0.1149 (0.1751)		0.0182 (0.0646)
R-Squared	0.012	0.012	0.005	0.006	0.007	0.007
N	3,695	3,695	3,020	3,020	3,699	3,699

* $p<0.1$; ** $p<0.05$; *** $p<0.01$
Standard errors in parentheses

Table 3.15

Robustness tests: Religious non-intense monitoring and earnings quality

The dependent variable in the first and second columns is the absolute value of discretionary accruals generated from the modified Jones model over 1998 – 2018. The dependent variable in the third and fourth columns is the absolute value of discretionary accruals generated from the modified Jones model augmented with a control for firm performance as in Kothari, Leone, and Wasley (2005). Religious non-monitoring-intensive board equals one when, on a non-monitoring-intensive board, a majority of the straddling from a base (or initial) principal monitoring committee to another principal monitoring committee is done by the religious independent directors, zero when, on a non-monitoring-intensive board, a majority of the said straddling is done by the non-religious independent directors. Externally busy board equals one when a majority of independent directors serve on three or more boards. Board size is the natural log of the number of directors. Board independence is the percentage of directors that are unaffiliated with the firm beyond their directorship. Audit committee independence is the percentage of independent directors on the audit committee. Firm size is the natural log of market value of equity. Book/market is the book value of equity divided by the market value of equity. The absolute change in net income is the absolute value of the change in net income between years $t-1$ and t. Leverage is the ratio of total assets to liabilities. Numbers in parentheses are p-values based on robust standard errors clustered at the firm level. Levels of significance are indicated by ***, **, and * for 1%, 5%, and 10%, respectively.

	Abnormal accruals		*ROA augmented abnormal accruals*	
	(1)	*(2)*	*(3)*	*(4)*
Religious Non-Monitoring-intensive board	0.0813 (0.0857)	0.0810 (0.0846)	0.0821 (0.0860)	0.0818 (0.0849)
Externally busy board	-0.0011 (0.0372)	0.0029 (0.0355)	-0.0014 (0.0373)	0.0026 (0.0356)
Board size	-0.1091 (0.0435)**	-0.1035 (0.0402)**	-0.1090 (0.0435)**	-0.1034 (0.0402)**
Board independence	0.0036 (0.0337)	0.0031 (0.0333)	0.0040 (0.0337)	0.0034 (0.0334)
Audit committee independence	-0.1566 (0.1599)	-0.1429 (0.1557)	-0.1589 (0.1610)	-0.1453 (0.1567)
Firm size	-0.0109 (0.0027)***	-0.0109 (0.0027)***	-0.0108 (0.0027)***	-0.0109 (0.0027)***
Book/market	0.0001 (0.0000)***	0.0001 (0.0000)***	0.0001 (0.0000)***	0.0001 (0.0000)***
Absolute change in net income	0.0000 (0.0000)***	0.0000 (0.0000)***	0.0000 (0.0000)***	0.0000 (0.0000)***
Leverage	-0.0000 (0.0000)	-0.0000 (0.0000)	-0.0000 (0.0000)	-0.0000 (0.0000)
Religious CEO		0.0691 (0.0698)		0.0683 (0.0696)
R-Squared	0.044	0.047	0.043	0.046
N	2,497	2,497	2,495	2,495

* p<0.1; ** p<0.05; *** p<0.01
Standard errors in parentheses

Table 3.16

Robustness tests: Religious non-intense monitoring and excess CEO compensation

The dependent variables in all columns are the actual residuals from the respective Table 3.8, Panel A regressions. Religious non-monitoring-intensive board equals one when, on a non-monitoring-intensive board, a majority of the straddling from a base (or initial) principal monitoring committee to another principal monitoring committee is done by the religious independent directors, zero when, on a non-monitoring-intensive board, a majority of the said straddling is done by the non-religious independent directors. Board size is the natural log of the number of directors. CEO duality equals one when the CEO also serves as board chair, zero otherwise. CEO directors is the number of directors who are CEOs of other firms. Religious CEO equals one if the CEO is religious, zero otherwise. Numbers in parentheses are p-values based on robust standard errors clustered at the firm level. Levels of significance are indicated by ***, **, and * for 1%, 5%, and 10%, respectively.

Excess CEO compensation

	Total Compensation		*Equity Compensation*		*Cash Compensation*	
	(1)	*(2)*	*(1)*	*(2)*	*(1)*	*(2)*
Religious Non-Monitoring-intensive board	-0.2175	-0.2171	0.0643	0.0660	-0.0834	-0.0836
	(0.1815)	(0.1815)	(0.2088)	(0.2091)	(0.1203)	(0.1204)
Board size	-0.0730	-0.0735	-0.2661	-0.2683	0.1048	0.1051
	(0.0866)	(0.0863)	(0.1146)**	(0.1139)**	(0.0832)	(0.0831)
CEO duality	0.0001	0.0002	0.0058	0.0060	0.0027	0.0026
	(0.0424)	(0.0425)	(0.0410)	(0.0410)	(0.0396)	(0.0396)
CEO directors	0.0605	0.0610	-0.0412	-0.0391	0.0143	0.0141
	(0.1460)	(0.1461)	(0.1339)	(0.1343)	(0.1296)	(0.1297)
Religious CEO		0.0139		0.0581		-0.0066
		(0.0819)		(0.1230)		(0.0647)
R-Squared	0.002	0.002	0.007	0.007	0.002	0.002
N	3,694	3,694	3,093	3,093	3,701	3,701

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Standard errors in parentheses

Table 3.17. Variables definitions

Board characteristics	
Directors	The number of directors on the board
Board size	The natural log of the number of directors
Independent directors	The number of independent directors on the board
Board independence	The percentage of directors that are unaffiliated with the firm beyond their directorship
Externally busy board	An indicator variable that equals 1 if a majority of independent directors serve on three or more boards, 0 otherwise
Audit committee membership	The number of directors on the audit committee
Compensation committee membership	The number of directors on the compensation committee
Nominating/governance committee membership	The number of directors on the nominating and governance committee
Audit committee independence	The percentage of independent directors on the audit committee
Compensation committee independence	The percentage of independent directors on the compensation committee
Nominating/governance committee independence	The percentage of independent directors on the nominating and governance committee
Monitoring-intensive board	An indicator variable that equals 1 if a majority of the independent directors are monitoring-intensive, 0 otherwise
Religious monitoring-intensive board	An indicator variable that equals 1 if, on a monitoring-intensive board, a majority of the straddling from a base (or initial) principal monitoring committee to another principal monitoring committee is done by the religious independent directors, 0 when, on a monitoring-intensive board, a majority of the said straddling is done by the non-religious independent directors
Religious non-monitoring-intensive board	An indicator variable that equals 1 if, on a non-monitoring-intensive board, a majority of the straddling from a base (or initial) principal monitoring committee to another principal monitoring committee is done by the religious independent directors, 0 when, on a non-monitoring-intensive board, a majority of the said straddling is done by the non-religious independent directors
Religious lead independent director	An indicator variable that equals 1 if the lead independent director is religious, 0 otherwise
Religious audit chair	An indicator variable that equals 1 if the chairman of the audit committee is religious, 0 otherwise
Religious compensation chair	An indicator variable that equals 1 if the chairman of the compensation committee is religious, 0 otherwise
Religious nominating and governance chair	An indicator variable that equals 1 if the chairman of the nominating and governance committee is religious, 0 otherwise
Religious majority monitoring chair	An indicator variable that equals 1 if a majority of the chairmen of the principal monitoring committees are religious, 0 otherwise
Powerful religious monitoring-intensive board	An indicator variable that equals 1 if the board is religious monitoring-intensive, the lead independent director is religious, and a majority of the chairmen of the principal monitoring committees are religious, 0 if the board is non-religious monitoring-intensive, the lead independent director is non-religious, and a majority of the chairmen of the principal monitoring committees are non-religious

Firm characteristics	
Firm size (*all other regression models*)	The natural log of total assets
Firm size (*accrual model*)	The natural log of the market value of equity
Absolute change in net income	The absolute value of the change in between years t-1 and t
Loss	An indicator variable that equals 1 for firms with two or more consecutive years of negative income, 0 otherwise
Leverage	The ratio of total assets to liabilities
Book/market	The book value of equity divided by the market value of equity
Stock return	The annual stock return less same-period return on the CRSP value-weighted portfolio of NYSE/Amex/Nasdaq stocks
ROA	Operating income before depreciation divided by total assets
SDRet	The standard deviations of stock return over the preceding five years
SDROA	The standard deviations of ROA over the preceding five years
1-yr Market-adjusted return	The annual stock return less same period return on the CRSP value-weighted portfolio of NYSE/Amex/Nasdaq stocks
2-yr Market-adjusted return	The bi-annual stock return less same period return on the CRSP value-weighted portfolio of NYSE/Amex/Nasdaq stocks
Institutional ownership	The percentage of outstanding shares owned by institutional investors
Abnormal accruals	The absolute value of discretionary accruals generated from the modified Jones model over 1998 – 2018
ROA-augmented abnormal accruals	The absolute value of discretionary accruals generated from a similar model augmented with a control for firm performance as in Kothari, Leone, and Wasley (2005)

CEO characteristics	
Religious CEO	An indicator variable that equals 1 if the CEO is religious*, zero otherwise
CEO age	The age of the CEO in years
CEO duality	An indicator variable that equals 1 if the CEO also serves as board chair, zero otherwise
CEO directors	The number of directors who are CEOs of other firms
CEO ownership	The proportion of outstanding shares beneficially owned by the CEO
Forced turnover	An indicator variable that equals 1 for firm-years with forced** CEO turnovers, 0 otherwise
All turnover	An indicator variable that equals 1 for firm-years with any CEO turnovers, 0 otherwise
Total compensation	The natural log of the sum of salary, bonus, the value of stock options and restricted stock granted during the year, long-term incentive payouts, and other miscellaneous compensation amounts
Equity compensation	The natural log of (one plus) the value of stock options and restricted stock awarded during the year
Cash compensation	The natural log of salary plus cash bonus
Total compensation (Excess)	The residuals from the regression of total compensation on the economic determinants in Table 8 (i.e., firm size, Book/market, stock return, ROA, SDRet, SDROA)
Equity compensation (Excess)	The residuals from the regression of equity compensation on the economic determinants in Table 8 (i.e., firm size, Book/market, stock return, ROA, SDRet, SDROA)
Cash compensation (Excess)	The residuals from the regression of cash compensation on the economic determinants in Table 8 (i.e., firm size, Book/market, stock return, ROA, SDRet, SDROA)

* *A CEO or a director is religious if s/he attended a religious-affiliated academic institution, non-religious otherwise*

** *A CEO turnover is considered forced if it was involuntary as in Peters and Wagner (Journal of Finance, vol. 69, no. 4, pp. 1529-1563, 2014)*

Appendix 3.IA (Designated Internet Appendix)
Examining related FHH results over a longer time horizon

As previously indicated, this chapter directly builds on Faleye, Hoitash, and Hoitash (2011) – hereafter FHH. Nevertheless, whereas FHH used a data sample that spans the period from 1998 to 2006, our base sample runs from 1998 to 2018 (i.e., some additional 12 years; and much further from the introduction of the Sarbanes-Oxley Act, or SOX, in 2002) and over 83% of our firm-years come from the period post-2006. Also, whereas FHH compose their performance measures based only on a 1-year holding period, we further evaluate the progression of the sensitivity of CEO turnover to firm performance as the performance measurement horizon shifts to a term longer than one year. Hence, empirically, we proceed in two main steps. In the first step, we evaluate the evolution of the actual FHH results over our longer sample and performance horizons. In the second step, we switch to the central thesis of this chapter. This Appendix details that first step.

A.1 Monitoring-Intensive Boards, Performance Horizons, and CEO Turnover

We begin by taking advantage of our longer sample data (spanning the 21 years from 1998 to 2018) to examine the stability and/or evolution of the related FHH results.[79] Table A.1 presents our logistic regression model for this purpose. The dependent variable equals one for forced turnovers and zero otherwise (in columns 1 and 2) and equals one for all turnovers and zero otherwise (in columns 3 and 4). Similar to FHH, our specific variable of interest at this point is the interaction term between monitoring intensity and 1-year firm performance; and consistent with the improved monitoring hypothesis[80], we find a negative coefficient, which is significant at the 1% level. Nevertheless, to understand the evolution of the actual probabilities of forced CEO turnover as the firm performance deteriorates, we compute the predictive margins[81] for the monitoring-intensive and non-monitoring-intensive boards. Figure A.1 graphs the predictive margins with the associated 95% confidence intervals (CIs). The vertical axis shows the actual probabilities of forced

[79] On this subject, FHH find that, compared to non-monitoring-intensive boards, monitoring-intensive boards exhibit a greater sensitivity of CEO turnover to [1-year] firm performance.
[80] See the main chapter; FHH.
[81] Predictive margins are statistics calculated from predictions of a previously fit model at fixed values of some covariates and averaging or otherwise integrating over the remaining covariates (e.g., see Buis, 2010; Hosmer, Lemeshow, and Sturdivant, 2013; Long and Freese, 2014; Norton, Wang, and Ai, 2004; Williams, 2012).

turnover. The horizontal axis shows firm performance[82] (or more aptly, in this case, underperformance). The latter is calculated over a 12-month or 1-year measurement horizon and ranges from -20% to 0%. Consequently, the graphs of the predictive margins of the monitoring-intensive and non-monitoring-intensive boards appropriately depict their respective sensitivities of CEO forced turnover to market-underperformance.

Generally consistent with FHH, we observe that monitoring-intensive boards indeed exhibit a greater sensitivity of CEO turnover to 1-year firm performance (i.e., performance over a 12-month measurement horizon). However, we also observe some interesting statistical nuances. For example, over the 1-year holding period, both board-types do not appear to be inclined to force a CEO out for matching or for barely underperforming the market up to -5%. Below this level of market-underperformance, nonetheless, we observe a significant statistical divergence. Specifically, while the non-monitoring-intensive boards remain steadily disinclined to forcing the CEO out, the probability of forced turnover rises exponentially for the monitoring-intensive boards, from 0 at 0%, to close to 0.5 at -20% (with a 95% CI between a little less than 0.2 and 0.8).

Returning to our regression results in Table A.1, we observe negative coefficients for CEO duality that are all significant at 1% both for forced-turnover (columns 1 and 2) and all turnovers (columns 3 and 4). This implies that even in the event of forced turnover it is significantly difficult to force a CEO that also chairs the board out. Overall, we observe the same qualitative results whether in consideration of forced turnovers only or all turnovers as a whole. However, we find that CEO ownership matters far less in the event of a forced turnover (perhaps sensibly so[83: a, b, c, d, e]) than in the event of voluntary turnover.

---

[82] For precision and/or emphasis, we systematically define firm performance as the market-adjusted stock return (i.e., the difference between a firm's stock return and that of the market. Both returns are calculated over the same holding period and the market is the CRSP value-weighted portfolio of the NYSE/Amex/Nasdaq stocks).

[83] (a). In 2013, Groupon co-founder and former CEO Andrew Mason was fired from the daily deals website four and a half years after its founding. Under Mason's leadership, the company's shares plummeted and the business faced serious financial challenges. Mason took responsibility for the company's poor financial state and after being fired, stated his employees: "After four and a half intense and wonderful years as CEO of Groupon, I've decided that I'd like to spend more time with my family. Just kidding -- I was fired today."
(b). Throughout 2017, the world watched as more and more Uber controversies unraveled. From sexual harassment allegations to self-driving car crashes and lawsuits with Google, the list goes on. Cultivating a company culture known for frat boys and out-of-control parties, Kalanick was asked to resign in 2017 so the company could begin repairing its damaged reputation.
(c). In 2006, Jack Dorsey, Evan Williams, Biz Stone, and Noah Glass launched Twitter. In the company's early days, Dorsey took the title of CEO, but it wasn't before long that discussion surfaced about his poor

Next, keeping all other elements of the previous regression model unchanged, we examine the progression of these sensitivities of CEO turnover to firm performance, when the performance measurement horizon is 24 months or 2 years. Table A.2 presents the associated logistic regression model. Again, the dependent variable equals one for forced turnovers and zero otherwise (in columns 1 and 2) and equals one for all turnovers and zero otherwise (in columns 3 and 4). Similarly, our specific variable of interest at this point is the interaction term between monitoring intensity and 2-year performance. Remarkably, although we still find a negative coefficient, this coefficient is non-significant even at the 10% level. To provide some statistical nuances, we similarly compute the predictive margins. Figure A.2 graphs these margins. Qualitatively or pictorially, the monitoring-intensive boards still appear to exhibit a greater sensitivity of CEO turnover to firm performance, on average. However, statistically, the latter appearance is not exactly significant, especially considering the overlaps of the confidence intervals.

In effect, we find that over a 2-year holding period, both the non-monitoring-intensive and the monitoring-intensive boards become increasingly inclined to force a CEO out for underperforming the market below the -5% level. However, while on average, the sensitivity remains higher for monitoring-intensive boards as performance further deteriorates, the (albeit, more gently) increasing sensitivity of the non-monitoring-intensive boards over this holding period precludes the divergences in sensitivity between both board types from being statistically significant, as the CIs largely and systematically overlap, even up to the -20% market-underperformance point.

A.2 Monitoring-Intensive Boards and Excess CEO Pay

Panel B of Table A.3 presents regressions explaining excess compensation (defined as the residuals from the respective Panel A of Table A.3 regressions). In other words, the dependent variables in all columns are the actual residuals from the corresponding Panel

---

management style, lack of communication with board members, and inability to fix its constantly crashing servers. In 2008, Dorsey was fired by co-founder Evan Williams, who was the main investor of the company and the chairman.

(d). While he was always recognized for his wit and intelligence, Steve Jobs's short temper was also widely known -- enough so that it cost him his leadership role in the company. In the 1980s, Apple's board found Jobs too young and temperamental to run the company, so in response, Jobs brought in Pepsi executive John Sculley. However, after many differences and disagreements, Sculley in 1985 convinced the board to let Jobs go.

(e). See https://www.entrepreneur.com/slideshow/321360#1 for more stories. Dorsey and Jobs returned later.

A of Table A.3 regressions. As is systematically the case, monitoring-intensive board equals one when a majority of independent directors serve on at least two of the three principal monitoring committees. Externally busy board equals one when a majority of independent directors serve on three or more boards. Board size is the natural log of the number of directors. CEO duality equals one when the CEO also serves as board chair, zero otherwise. The CEO directors variable refers to the number of directors who are CEOs of other firms. Religious CEO equals one if the CEO is religious, zero otherwise.

First, similar to FHH, we find a negative association between monitoring intensity and excess equity compensation that is significant at the 5% level. We also find no significant association between monitoring intensity and excess cash compensation. However, whereas we effectively observe hints of a negative association between monitoring intensity and total CEO compensation, our results in columns 1 and 2 (Panel B of Table A.3) suggest, at the least, that for the additional 12-year period post-2006 (i.e., for the firm-years much further from the enactment of the SOX) contained within our data sample, although monitoring intensity alone continues to, on average, significantly reduce excess equity compensation, the reductions have not been huge enough to similarly impact excess total CEO compensation, all other things equal. Our results also show positive associations between all three excess compensation items and the external busyness of the board, all of which are significant at the 1% level. The latter suggests that that excess CEO compensation broadly exacerbates when directors are eternally busy.

A.3 Monitoring-Intensive Boards and Earnings Quality

Table A.4 presents the results of the associated regressions. The dependent variable in columns 1 and 2 is the absolute value of discretionary accruals generated from the modified Jones model using our longer horizon data sample (that spans the 21 years from 1998 to 2018). Precisely, we estimate the discretionary accrual model each year using all firm-year observations in the same two-digit SIC code as follows:

$$TA_{it} = \frac{\beta_0}{Assets_{it-1}} + \beta_1(\Delta Sales_{it} - \Delta Receivables_{it}) + \beta_2 PPE_{it} + \varepsilon_{it}$$

where TA_{it} is the total accruals scaled by lagged total assets; $\Delta Sales_{it}$ is the change in net sales scaled by lagged total assets; $\Delta Receivables_{it}$ is the change in net receivables scaled by lagged total assets; PPE_{it} is gross property, plant, and equipment scaled by the lagged total asset. The fitted value of the regression model is taken to be the non-discretionary accruals in an industry-year with the given level of sales changes, property, plant, and equipment. The regression residual is considered not dictated by the firm and industry conditions and is taken to be the discretionary component of accruals.

The dependent variable in columns 3 and 4 is the absolute value of discretionary accruals generated from the modified Jones model augmented with ROA (to control for firm performance, as in Kothari, Leone, and Wasley, 2005). Specifically, we estimate the ROA-augmented discretionary accrual model each year using all firm-year observations in the same two-digit SIC code as follows:

$$TA_{it} = \frac{\beta_0}{Assets_{it-1}} + \beta_1(\Delta Sales_{it} - \Delta Receivables_{it}) + \beta_2 PPE_{it} + \beta_3 ROA_{it-1} + \varepsilon_{it}$$

Where ROA_{it-1} is the lagged return on assets and is a proxy for firm performance. All other components are as previously defined (e.g., see Kothari, Leone, and Wasley, 2005; FHH; CKLP). Monitoring-intensive board, externally busy board, board size, and book/market are as previously defined (see Table 3.17 of the main chapter). Board independence is the percentage of directors that are unaffiliated with the firm beyond their directorship. Audit committee independence is the percentage of independent directors on the audit committee. Firm size, in this case, is the natural log of the market value of equity and the absolute change in net income is the absolute value of the change in net income between years $t - 1$ and t.

Consistent with FHH, the signs of the coefficients (for the monitoring-intensive board variable) are negative across all columns and suggest that monitoring-intensive boards continue to be associated with reductions in discretionary accruals. Although the coefficients appear only statistically significant at the 10% level, their magnitudes point to some highly economically significant improvements in earnings quality. For example, the smallest coefficient in column 4 suggests that as monitoring becomes more intense the ratio of discretionary accruals to total assets is lower by 3.28%. Given that the mean value of discretionary accruals in the subsample is about 11.5% of total assets, this effectively represents an economically significant 28.5% reduction in abnormal accruals, all other things equal. The results for other variables are also consistent with previous research. For example, we also find that earnings quality is higher among larger firms, increases with board size and independence, and with the independence of the audit committee (e.g., see Klein, 2002; Larcker, Richardson, and Tuna, 2007). Also, an infinitesimal positive coefficient that is statistically significant at the 1% level suggests that earnings quality will tend to be worse among firms with significant net income changes (e.g, see Klein 2002).

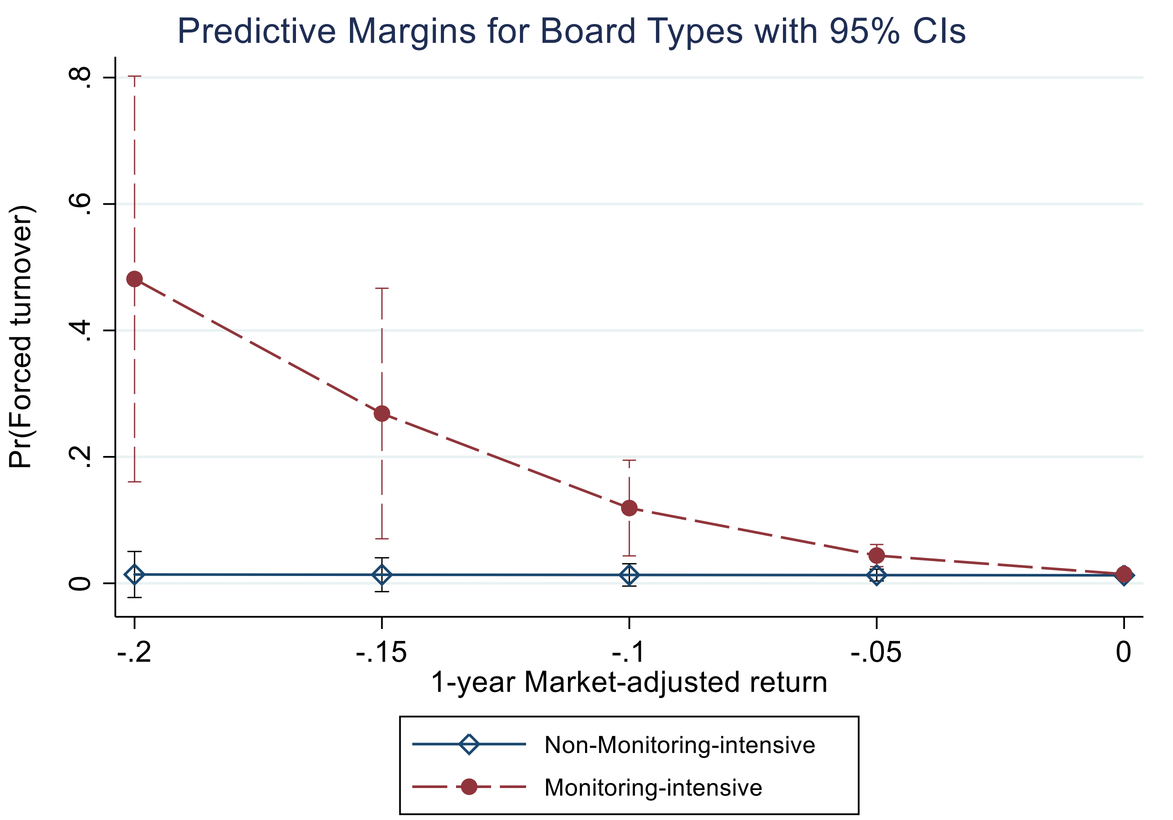


Figure A.1
Predictive margins of monitoring-intensive boards and 1-yr performance
Predictive margins are statistics calculated from predictions of a previously fit model at fixed values of some covariates and averaging or otherwise integrating over the remaining covariates (e.g., see Buis, 2010; Hosmer, Lemeshow, and Sturdivant, 2013; Long and Freese, 2014; Norton, Wang, and Ai, 2004; Williams, 2012). The previously fit model is the logistic regression model in Table 3.4, column 1. The vertical axis shows the actual probabilities of forced turnover. The horizontal axis shows the 1-year firm performance or market-adjusted return (defined as the annual stock return less same period return on the CRSP value-weighted portfolio of NYSE/Amex/Nasdaq stocks). The continuous navy blue line with a hollow diamond represents the average sensitivity of forced CEO turnover to 1-year firm performance for the non-monitoring-intensive boards. The dashed maroon line with solid maroon circles represents the average sensitivity of forced CEO turnover to 1-year firm performance for monitoring-intensive boards. Both lines have vertical outlines of the 95% CI.

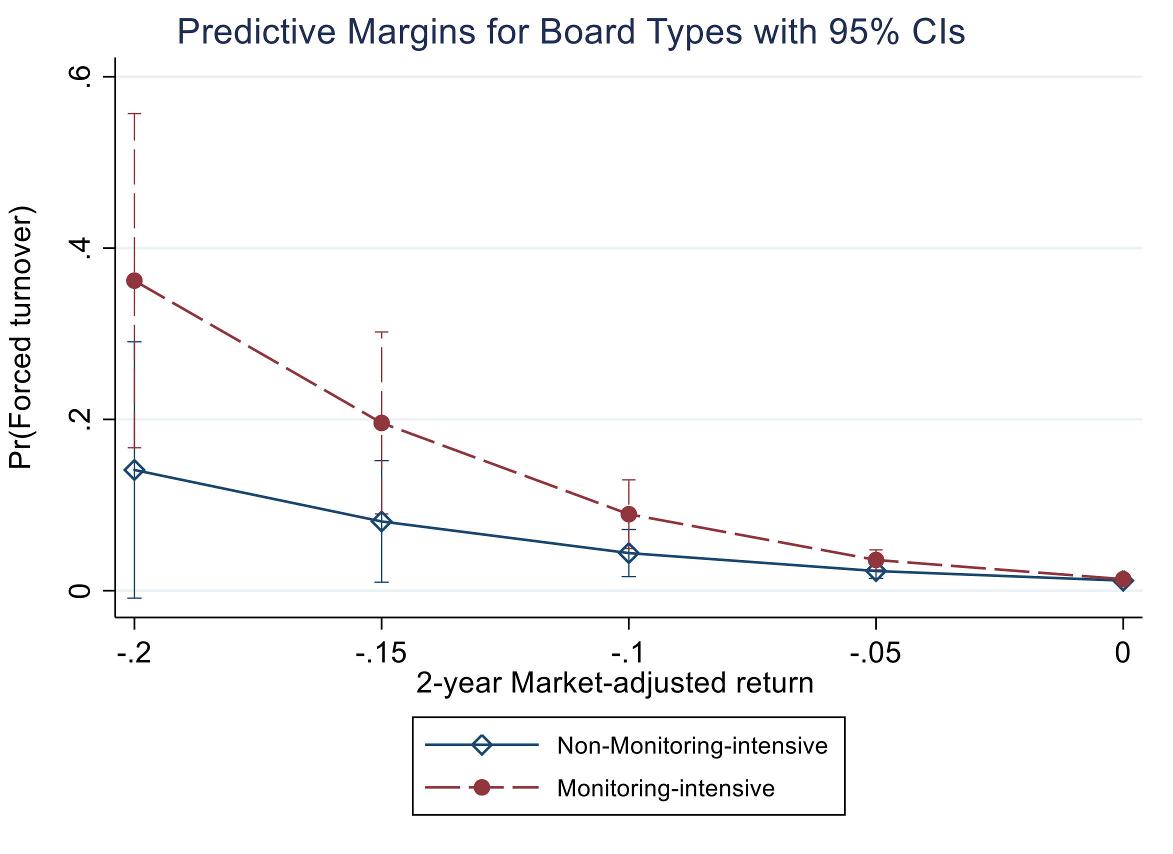


Figure A.2
Predictive margins of monitoring-intensive boards and 2-yr performance
Predictive margins are statistics calculated from predictions of a previously fit model at fixed values of some covariates and averaging or otherwise integrating over the remaining covariates (e.g., see Buis, 2010; Hosmer, Lemeshow, and Sturdivant, 2013; Long and Freese, 2014; Norton, Wang, and Ai, 2004; Williams, 2012). The previously fit model is the logistic regression model in Table 3.5, column 1. The vertical axis shows the actual probabilities of forced turnover. The horizontal axis shows the 2-year firm performance or market-adjusted return (defined as the bi-annual stock return less same period return on the CRSP value-weighted portfolio of NYSE/Amex/Nasdaq stocks). The continuous navy blue line with a hollow diamond represents the average sensitivity of forced CEO turnover to 2-year firm performance for the non-monitoring-intensive boards. The dashed maroon line with solid maroon circles represents the average sensitivity of forced CEO turnover to 2-year firm performance for monitoring-intensive boards. Both lines have vertical outlines of the 95% CI.

Table A.1
Monitoring intensity and sensitivity of CEO turnover to 1-year market-adjusted firm performance.

The dependent variable in columns 1 and 2 equals one for firm-years with forced CEO turnovers, zero otherwise. The dependent variable in columns 3 and 4 equals one for firm-years with any CEO turnovers, zero otherwise. Monitoring-intensive board equals one when a majority of independent directors serve on at least two of the three principal monitoring committees. 1-year market-adjusted return is annual stock return less same period return on the CRSP value-weighted portfolio of NYSE/Amex/Nasdaq stocks. CEO duality equals one when the CEO also serves as board chair, zero otherwise. Externally busy board equals one when a majority of independent directors serve on three or more boards, zero otherwise. Board independence equals one when a majority of directors are independent, zero otherwise. Board size is the natural log of the number of directors. Institutional ownership is the percentage of outstanding shares owned by institutional investors. CEO ownership is the proportion of outstanding shares beneficially owned by the CEO. Firm size is the natural log of total assets. CEO age is measured in years. Religious CEO equals one if the CEO is religious, zero otherwise. Each regression includes year and sector fixed effects. Numbers in parentheses are p-values based on robust standard errors clustered at the firm level. Levels of significance are indicated by ***, **, and * for 1%, 5%, and 10%, respectively.

	Forced Turnover		*All Turnover*	
	(1)	*(2)*	*(3)*	*(4)*
Monitoring-intensive board	0.1438 (0.2658)	0.1425 (0.2652)	0.0061 (0.0886)	0.0097 (0.0887)
1-yr Market-adjusted return	-0.4955 (6.8979)	-0.5051 (6.8883)	-4.5360 (2.3370)*	-4.4916 (2.3478)*
Monitoring-intensive board x 1-yr Market-adjusted return	-23.1064 (8.3921)***	-23.1401 (8.3587)***	-3.5959 (3.1030)	-3.6050 (3.1065)
CEO duality	-1.1077 (0.2566)***	-1.1077 (0.2567)***	-0.4828 (0.1006)***	-0.4824 (0.1006)***
Externally busy board	2.1007 (1.0728)*	2.0964 (1.0729)*	-0.6076 (0.7336)	-0.5976 (0.7335)
Board independence	1.0581 (0.9015)	1.0723 (0.8998)	0.6695 (0.3508)*	0.6678 (0.3515)*
Board size	1.1066 (0.5785)*	1.0993 (0.5831)*	0.2636 (0.2258)	0.2723 (0.2255)
Institutional ownership	-0.0742 (0.7067)	-0.0787 (0.7094)	0.1204 (0.2491)	0.1239 (0.2490)
CEO ownership	-0.0260 (0.0342)	-0.0262 (0.0344)	-0.0652 (0.0177)***	-0.0647 (0.0177)***
Firm size	0.0085 (0.0849)	0.0093 (0.0851)	-0.0172 (0.0288)	-0.0187 (0.0288)
CEO age	0.0560 (0.0155)***	0.0560 (0.0154)***	0.0700 (0.0068)***	0.0699 (0.0068)***
Religious CEO		-0.1431 (0.8074)		0.2128 (0.2289)
Year fixed effects	Yes	Yes	Yes	Yes
Sector fixed effects	Yes	Yes	Yes	Yes
Pseudo R-Squared	0.126	0.126	0.088	0.088
N	6,894	6,894	8,309	8,309

* $p<0.1$; ** $p<0.05$; *** $p<0.01$
Standard errors in parentheses

Table A.2
Monitoring intensity and sensitivity of CEO turnover to 2-year market-adjusted firm performance.

The dependent variable in columns 1 and 2 equals one for firm-years with forced CEO turnovers, zero otherwise. The dependent variable in columns 3 and 4 equals one for firm-years with any CEO turnovers, zero otherwise. Monitoring-intensive board equals one when a majority of independent directors serve on at least two of the three principal monitoring committees. 2-year market-adjusted return is bi-annual stock return less same period return on the CRSP value-weighted portfolio of NYSE/Amex/Nasdaq stocks. CEO duality equals one when the CEO also serves as board chair, zero otherwise. Externally busy board equals one when a majority of independent directors serve on three or more boards, zero otherwise. Board independence equals one when a majority of directors are independent, zero otherwise. Board size is the natural log of the number of directors. Institutional ownership is the percentage of outstanding shares owned by institutional investors. CEO ownership is the proportion of outstanding shares beneficially owned by the CEO. Firm size is the natural log of total assets. CEO age is measured in years. Religious CEO equals one if the CEO is religious, zero otherwise. Each regression includes year and sector fixed effects. Numbers in parentheses are *p*-values based on robust standard errors clustered at the firm level. Levels of significance are indicated by ***, **, and * for 1%, 5%, and 10%, respectively.

	Forced Turnover		*All Turnover*	
	(1)	*(2)*	*(3)*	*(4)*
Monitoring-intensive board	0.1381 (0.2849)	0.1354 (0.2838)	0.0208 (0.0895)	0.0241 (0.0896)
2-yr Market-adjusted return	-13.9731 (3.9036)***	-13.9298 (3.8862)***	-7.0083 (1.6363)***	-7.0010 (1.6400)***
Monitoring-intensive board x 2-yr Market-adjusted return	-6.8049 (4.7103)	-6.9497 (4.7044)	-0.4538 (2.2412)	-0.4177 (2.2418)
CEO duality	-1.0942 (0.2603)***	-1.0960 (0.2606)***	-0.4911 (0.1025)***	-0.4903 (0.1024)***
Externally busy board	2.0802 (1.0541)**	2.0738 (1.0538)**	-0.5876 (0.7459)	-0.5791 (0.7455)
Board independence	0.7398 (0.8810)	0.7744 (0.8783)	0.8238 (0.3681)**	0.8195 (0.3686)**
Board size	1.0841 (0.5748)*	1.0667 (0.5796)*	0.2729 (0.2256)	0.2812 (0.2252)
Institutional ownership	0.2628 (0.6988)	0.2544 (0.7000)	0.2348 (0.2564)	0.2374 (0.2562)
CEO ownership	-0.0245 (0.0334)	-0.0247 (0.0335)	-0.0656 (0.0187)***	-0.0652 (0.0186)***
Firm size	0.0313 (0.0905)	0.0334 (0.0909)	-0.0107 (0.0291)	-0.0122 (0.0292)
CEO age	0.0553 (0.0156)***	0.0553 (0.0156)***	0.0698 (0.0069)***	0.0697 (0.0069)***
Religious CEO		-0.3204 (0.8540)		0.1998 (0.2368)
Year fixed effects	Yes	Yes	Yes	Yes
Sector fixed effects	Yes	Yes	Yes	Yes
Pseudo R-Squared	0.152	0.152	0.094	0.094
N	6,734	6,734	8,124	8,124

* $p<0.1$; ** $p<0.05$; *** $p<0.01$
Standard errors in parentheses

Table A.3

Monitoring intensity and excess CEO compensation.

Panel A presents regressions predicting normal CEO compensation as a function of the economic determinants of executive pay during 1998–2018. Total compensation is the natural log of the sum of salary, bonus, the value of stock options and restricted stock granted during the year, long-term incentive payouts, and other miscellaneous compensation amounts. Equity compensation is the natural log of (one plus) the value of stock options and restricted stock awarded during the year. Cash compensation is the natural log of salary plus cash bonus. Firm size is the natural log of total assets. Book/market is the book value of equity divided by the market value of equity. Stock return is the annual stock return less same-period return on the CRSP value-weighted portfolio of NYSE/Amex/Nasdaq stocks. ROA is operating income before depreciation divided by total assets. SDRet and SDROA are the respective standard deviations of stock return and ROA over the preceding five years. Each regression includes year and sector fixed effects.

Panel B presents regressions explaining excess compensation, defined as the residuals from the respective Panel A regressions. The dependent variables in all columns are the actual residuals. Monitoring-intensive board equals one when a majority of independent directors serve on at least two of the three principal monitoring committees. Externally busy board equals one when a majority of independent directors serve on three or more boards. Board size is the natural log of the number of directors. CEO duality equals one when the CEO also serves as board chair, zero otherwise. CEO directors is the number of directors who are CEOs of other firms. Religious CEO equals one if the CEO is religious, zero otherwise. Numbers in parentheses are p-values based on robust standard errors clustered at the firm level. Levels of significance are indicated by ***, **, and * for 1%, 5%, and 10%, respectively.

Panel A: Economic determinants

	Total Compensation	*Equity Compensation*	*Cash Compensation*
Firm size	0.4057 (0.0176)***	0.5054 (0.0157)***	0.1607 (0.0148)***
Book/market	-0.1735 (0.0612)***	-0.1938 (0.0717)***	-0.0572 (0.0330)*
Stock return	0.2307 (0.4093)	-0.3681 (0.4158)	-0.2230 (0.3520)
ROA	0.4180 (0.1578)***	0.5158 (0.2261)**	0.1921 (0.1495)
SDRet	0.2764 (0.3130)	0.9382 (0.4382)**	-0.3363 (0.2461)
SDROA	0.8105 (1.0557)	2.1660 (1.3146)*	-1.8941 (1.0803)*
Year fixed effects	Yes	Yes	Yes
Sector fixed effects	Yes	Yes	Yes
R-Squared	0.517	0.565	0.259
N	8,058	6,632	8,072

Panel B: Excess compensation

	Total Compensation		*Equity Compensation*		*Cash Compensation*	
	(1)	*(2)*	*(1)*	*(2)*	*(1)*	*(2)*
Monitoring-intensive board	-0.0397 (0.0302)	-0.0391 (0.0302)	-0.0803 (0.0368)**	-0.0793 (0.0367)**	0.0101 (0.0269)	0.0104 (0.0270)
Externally busy board	0.7440 (0.0353)***	0.7458 (0.0355)***	0.7840 (0.0387)***	0.7866 (0.0389)***	0.2610 (0.0282)***	0.2617 (0.0286)***
Board size	0.1362 (0.0677)**	0.1364 (0.0676)**	0.0253 (0.0935)	0.0261 (0.0935)	0.1360 (0.0730)*	0.1361 (0.0730)*
CEO duality	0.0382 (0.0329)	0.0384 (0.0329)	-0.0172 (0.0378)	-0.0171 (0.0377)	0.0340 (0.0299)	0.0341 (0.0298)
CEO directors	0.0363 (0.0820)	0.0373 (0.0820)	0.0133 (0.1052)	0.0147 (0.1053)	-0.0016 (0.0771)	-0.0012 (0.0772)
Religious CEO		0.0443 (0.0686)		0.0642 (0.1064)		0.0176 (0.0464)
R-Squared	0.005	0.005	0.003	0.004	0.004	0.004
N	8,058	8,058	6,632	6,632	8,072	8,072

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

Standard errors in parentheses

Table A.4

Monitoring intensity and earnings quality

The dependent variable in the first and second columns is the absolute value of discretionary accruals generated from the modified Jones model over 1998 – 2018. The dependent variable in the third and fourth columns is the absolute value of discretionary accruals generated from the modified Jones model augmented with a control for firm performance as in Kothari, Leone, and Wasley (2005). Monitoring-intensive board equals one when a majority of independent directors serve on at least two of the three principal monitoring committees. Externally busy board equals one when a majority of independent directors serve on three or more boards. Board size is the natural log of the number of directors. Board independence is the percentage of directors that are unaffiliated with the firm beyond their directorship. Audit committee independence is the percentage of independent directors on the audit committee. Firm size is the natural log of market value of equity. Book/market is the book value of equity divided by the market value of equity. The absolute change in net income is the absolute value of the change in net income between years $t-1$ and t. Leverage is the ratio of total assets to liabilities. Numbers in parentheses are p-values based on robust standard errors clustered at the firm level. Levels of significance are indicated by ***, **, and * for 1%, 5%, and 10%, respectively.

	Abnormal accruals		*ROA augmented abnormal accruals*	
	(1)	*(2)*	*(3)*	*(4)*
Monitoring-intensive board	-0.0335 (0.0175)*	-0.0331 (0.0177)*	-0.0332 (0.0176)*	-0.0328 (0.0177)*
Externally busy board	-0.0182 (0.0581)	-0.0179 (0.0583)	-0.0196 (0.0584)	-0.0193 (0.0585)
Board size	-0.1914 (0.0482)***	-0.1911 (0.0485)***	-0.1912 (0.0481)***	-0.1909 (0.0485)***
Board independence	-0.0317 (0.0420)	-0.0318 (0.0420)	-0.0322 (0.0422)	-0.0323 (0.0422)
Audit committee independence	-0.3448 (0.1600)**	-0.3444 (0.1604)**	-0.3500 (0.1598)**	-0.3496 (0.1602)**
Firm size	-0.0205 (0.0044)***	-0.0205 (0.0044)***	-0.0206 (0.0045)***	-0.0206 (0.0045)***
Book/market	-0.0002 (0.0002)	-0.0002 (0.0002)	-0.0002 (0.0002)	-0.0002 (0.0002)
Absolute change in net income	0.0000 (0.0000)***	0.0000 (0.0000)***	0.0000 (0.0000)***	0.0000 (0.0000)***
Leverage	-0.0000 (0.0000)	-0.0000 (0.0000)	-0.0000 (0.0000)	-0.0000 (0.0000)
Religious CEO		0.0117 (0.0483)		0.0122 (0.0488)
R-Squared	0.084	0.084	0.084	0.084
N	7,011	7,011	7,006	7,006

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

Standard errors in parentheses

References

Acharya, Viral V., Heitor Almeida, and Malcolm Baker, 2015, Introduction: New perspectives on corporate capital structures, *Journal of Financial Economics* 118, 551–552.

Ackerman, A., 2021, NASDAQ's plan for Diversity on Corporate Boards is decried by Senate Republicans, *The Wall Street Journal* (Dow Jones & Company).

Adams, R.B., 2009. Asking directors about their dual roles. *Unpublished Working Paper*, University of Queensland.

Adams, Renée B., Heitor Almeida, and Daniel Ferreira, 2005, Powerful CEOs and their impact on corporate performance, *Review of Financial Studies* 18, 1403–1432.

Adams, Renée B., and Daniel Ferreira, 2007, A theory of Friendly Boards, *The Journal of Finance* 62, 217–250.

Adams, R.B., Ferreira, D., 2009. Women in the boardroom and their impact on governance and performance. *Journal of Financial Economics* 94, 291–309.

Adams, R. B., and Funk, P., 2012, Beyond the glass ceiling: Does gender matter? *Management Science 58, 219–235*.

Aggarwal, R., Drake, P. P., Kobor, A. and Noronha, G. (2010). Capital Structure. *CFA Institute*. ISBN 978-1-946442-0

Agrawal, A., and Knoeber, C., 2001. Do some outside directors play a political role? Journal of Law and Economics 44, 179–198.

Ahern, K.R., Dittmar, A.K., 2012. The changing of the boards: The impact on firm valuation of mandated female board representation. *Quarterly Journal of Economics* 127, 137–197.

Akerlof, George, and Rachel E. Kranton, "Economics and Identity", *Quarterly Journal of Economics* 115-3, August 2000, pp. 715-753.

Akerlof, George, and Rachel E. Kranton, "Identity and the Economics of Organizations" *Journal of Economic Perspectives* 19-1, Winter 2005, 9-32

Albuquerque, R. A., and Miao, J. (2013). CEO Power, Compensation and Governance. Annals of Economics and Finance, Vol. 14 (2), pp. 417-452

Allport, Gordon, 1954, The Nature of Prejudice, New York: Addison Wesley.

Almazan, Andres, and Javier Suarez, 2003, Entrenchment and severance pay in optimal governance structures, Journal of Finance 58, 519– 547.

Andersen, S., Harrison, G.W., ... Rutström, E.E., 2008. Eliciting risk and time preferences. *Econometrica* 76, 583–618.

Ang, B., 2019. Board Diversity by US Region. *Harvard Law School Forum on Corporate Governance and Financial Regulation*. Saturday, February 16, 2019.

Asquith, Paul, Robert Gertner, and David Scharfstein, 1994, Anatomy of financial distress: An examination of junk bond issuers, Quarterly Journal of Economics 109, 625–658.

Baker, Malcolm, and Jeffrey Wurgler, 2002, Market timing and capital structure, *The Journal of Finance* 57, 1–32.

Baker, Andrew, David F. Larcker, and Charles C. Wang, 2021, How much should we trust staggered difference-in-differences estimates?, *SSRN Electronic Journal*.

Baranchuk, N., Dybvig, P.H., 2009. Consensus in diverse corporate boards. *Review of Financial Studies* 22, 715–747.

Barnett, T. , Bass, K. , Brown, G. , 1996. Religiosity, ethical ideology and intentions to report a peer's wrongdoing. *J. Bus. Ethics* 15 (11), 1161–1175 .

Barro, R. , McCleary, R., 2003. Religion and economic growth. *Am. Sociol. Rev.* 68, 760–781 .

Barsky, R. , Juster, F. , Kimball, M. , Shapiro, M. , 1997. Preference parameters and behavioral heterogeneity: an experimental approach in the health and retirement study. *Quart. J. Econ.* 112, 1937–1996 .

Bartlett, Robert, and Frank Partnoy, 2020, The misuse of Tobin's q, *Vanderbilt Law Review*, Volume 73, Number 2.

Bebchuk, L.A., Cohen, A., 2005. The costs of entrenched boards. *Journal of Financial Economics* 78, 409–433.

Bebchuk, L.A., Fried, J., 2004. Pay without Performance: The Unfulfilled Promise of Executive Compensation. *Harvard University Press*, Cam bridge, MA.

Bebchuk, Lucian.A., and J.M. Fried, 2010, "Paying for Long-Term Performance." University of Pennsylvania Law Review 158, 1915-1960

Becker, G. , 1957. The Economics of Discrimination. *University of Chicago Press Economics Books*, Chicago.

Beetsma, Roel and Peter Schotman, 2001. Measuring risk attitudes in a natural experiment: Data from the television game show Lingo, *Economic Journal* 111, 821–848.

Bénabou, Roland and Jean Tirole, "Identity, Morals, and Taboos: Beliefs as Assets" *Quarterly Journal of Economics* 126(2), 2011, pp. 805-855.

Bénabou, Roland and Jean Tirole, "Individual and Corporate Social Responsibility" *Economica* 77(305), January 2010, pp. 1–19.

Bénabou, Roland and Jean Tirole, "Intrinsic and Extrinsic Motivation" *Review of Economic Studies* 70(3), 2003, pp. 489–520.

Bennhold, K., 2009. Where Would We Be If Women Ran Wall Street? *NY. TIMES*, Feb. 1, 2009.

Berliant, Marcus, and Masahisa Fujita, 2012, Culture and diversity in knowledge creation, *Regional Science and Urban Economics* 42, 648–662.

Bernile, G., Bhagwat, V., Yonker, S., 2018. Board diversity, firm risk, and corporate policies. *Journal of Financial Economics* 127, 588–612.

Bertrand, M. and Mullainathan, S. (2003). Enjoying the Quiet Life? Corporate Governance and

Managerial Preferences. *Journal of Political Economy*, 2003, vol. 111, no. 5
Bertrand, M., and A. Schoar, 2003, Managing with style: The effect of managers on firm policies, *The Quarterly Journal of Economics* 118, 1169–1208.
Billings, Mary Brooke, April Klein, and Yanting Shi, 2021, Investors' response to the #METOO Movement: Does corporate culture matter?, *SSRN Electronic Journal*.
Black, Sandra E., Paul J. Devereux, Petter Lundborg, and Kaveh Majlesi, 2017, On the origins of risk-taking in financial markets, *The Journal of Finance* 72, 2229–2278
Blalock, Hubert, 1967, Toward a Theory of Minority-Group Relations, New York: Wiley.
Blau, Peter, 1977, Inequality and Heterogeneity. New York: Free Press.
Bliss, R.T., Potter, M.E., Schwarz, C., 2012. Decision making and risk aversion in the Cash Cab. *Journal of Economic Behavior and Organization* 84, 163–173.
Bogan, V.L., Just, D.R., Dev, C.S., 2013. Team gender diversity and investment decision-making behavior. *Review of Behavioral Finance* 5, 134–152.
Boone, A., Field, L., Karpoff, J., Raheja, C., 2007, The determinants of corporate board size and composition: an empirical analysis. Journal of Financial Economics 85, 66–101.
Boquist, J. A, Milbourn, T, and Thakor, A. 2010. Value Sphere, The: The Corporate Executives' Handbook For Creating And Retatining Shareholder Wealth... 4th ed. *World Scientific*
Bradley, Michael, Gregg, Jarrell, and E. Han Kim, 1984, On the existence of an optimal capital structure: Theory and evidence, Journal of Finance 39, 857– 878.
Bromiley, Philip, 1991, Testing a causal model of corporate risk taking and performance, *Academy of Management Journal* 34, 37–59.
Browning, P. C. and Sparks, W. L. (2015). The Director's Manual: A Framework for Board Governance. *John Wiley & Sons, Inc.*, Hoboken, New Jersey
Buis, M. L. (2010). Stata Tip 87: Interpretation of interactions in nonlinear models. *The Stata Journal*, 10(2), pages 305-308.
Bunderson, S. and Thakor, A. V. 2020. Personal and organizational higher purpose : survey results, *Working paper*
Cai, Y., Kim, Y., Li, S., Pan, C., 2019. Tone at the Top: CEOs' Religious Beliefs and Earnings Management, *Journal of Banking and Finance* 106, 195–213
Calder-Wang, Sophie, Paul Gompers, and Kevin Huang, 2021, Diversity and performance in entrepreneurial teams, *SSRN Electronic Journal*.
Carter, D., Simkins, B., Simpson, W., 2003, Corporate governance, board diversity, and firm value. Financial Review 38, 33–53.
Carter, D.A., D'Souza, F., ... Simpson, W.G., 2010. The gender and ethnic diversity of US boards and board committees and firm financial performance. *Corporate Governance: An International Review* 18, 396–414.
Castilo, M., Leo, G., Petrie, R. Room Composition Effects and Risk Taking by Gender.

Experimental Economics, Forthcoming

Cengiz, D., A. Dube, A. Lindner, and B. Zipperer (2019, 8). The Effect of Minimum Wages on Low-Wage Jobs*. The Quarterly Journal of Economics 134 (3), 1405–1454.

Cesarini, David, Magnus Johannesson, Paul Lichtenstein, Örjan Sandewall, and Björn Wallace, 2010, Genetic variation in financial decision-making, *The Journal of Finance* 65, 1725–1754

Chen, H. , Huang, H. , Lobo, G. , Wang, C. , 2016. Religiosity and the cost of debt. *J. Banking and Finance*. 70, 70–85 .

Chen, J., W.S. Leung, W. Song, and M. Goergen. 2019. Why Female Board Representation Matters: The Role of Female Directors in Reducing Male CEO Overconfidence. Journal of Empirical Finance, 53, 70-90.

Coles, J., Daniel, N., and Naveen, L., 2008, Boards: does one size fit all? Journal of Financial Economics 87, 329–356.

Core, J.E., Holthausen, R.W., Larcker, D.F., 1999. Corporate governance, chief executive officer compensation, and firm performance. *Journal of Financial Economics* 51, 371–406.

Coughlan, A.T., Schmidt, R.M., 1985. Executive compensation, management turnover, and firm performance: an empirical investigation. Journal of Accounting and Economics 7, 43–66.

Cranney, S. 2013. Do People Who Believe in God Report More Meaning in Their Lives? The Existential Effects of Belief. *Journal for the Scientifc Study of Religion*

Creary, S. J., McDonnel, M-H., Ghai, S., and Scruggs, J., 2019. When and Why Diversity Improves Your Board's Performance, *Harvard Business Review* (online)

Cuotu, D., 2014. *Lessons in power: Lyndon Johnson revealed.* Harvard Business Review (online).

Dechow, P.M., Sloan, R., Sweeney, A., 1995. Detecting earnings management. The Accounting Review 70, 193–225.

DeFond, M.L., Jiambalvo, J., 1994. Debt covenant violation and manipulation of accruals. Journal of Accounting and Economics 17, 145–176.

Demsetz, H., Lehn, K., 1985. The structure of corporate ownership: causes and consequences. Journal of Political Economy 93, 1155–1177.

Denis, D.J., Denis, D.K., Sarin, A., 1997. Ownership structure and top executive turnover. Journal of Financial Economics 45, 193–221.

Diaz, J. , 2000. Religion and gambling in Sin–city: a statistical analysis of the relationship between religion and gambling patterns in Las Vegas residents. *Soc. Sci. J.* 37, 453–458 .

Eckbo, B. Espen and Nygaard, Knut and Thorburn, Karin S., Board Gender-Balancing and Firm Value (February 24, 2019). *European Corporate Governance Institute (ECGI) - Finance Working Paper* No. 463/2016.

Eckel, C. C., and Grossman, P. J., 2002. "Sex differences and statistical stereotyping in attitudes toward financial risk." *Evolution and Human Behavior 23(4): 281-295.*

Eckel, C. C., and Grossman, P. J., 2008. "Forecasting risk attitudes: An experimental study of actual and forecast risk attitudes of women and men." *Journal of Economic Behavior and Organization 68(1): 1-17.*

Edmans, A., Xavier, G., and Augustin L., 2009. "A multiplicative model of optimal CEO incentives in market equilibrium." The Review of Financial Studies 22, no. 12: 4881-4917.

Erhardt, N.L., Werbel, J.D., Shrader, C.B., 2003. Board of Director Diversity and Firm Financial Performance. *Corporate Governance* 11, 102–111.

Faccio, M., Marchica, M.T., Mura, R., 2016. CEO gender, corporate risk-taking, and the efficiency of capital allocation. *Journal of Corporate Finance* 39, 193–209.

Faleye, O., Hoitash, R., Hoitash, U., 2011. The Costs of Intense Board Monitoring *Journal of Financial Economics* 101, 160–181.

Fama, E. F., 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance, 25(2), 383–417.*

Fama, Eugene, and Kenneth French, 1992, The cross-section of expected returns, Journal of Finance 46, 427– 466.

Ferris, S.P., Jagannathan, M., and Pritchard, A.C ,2003. Too busy to mind the business? Monitoring by directors with multiple board appointments. *Journal of Finance* 58, 1087–1111.

Fich, E.M., and Shivdasani, A., 2006. Are busy boards effective monitors. *Journal of Finance* 61, 689–724.

Fischer, Edwin O., Robert Heinkel, and Josef Zechner, 1989, Optimal dynamic capital structure choice: Theory and tests, Journal of Finance 44, 19–40.

Forbes, D. P., & Milliken, F. J., 1999. Cognition and Corporate Governance: Understanding Boards of Directors as Strategic Decision-Making Groups. *The Academy of Management Review*, *24*(3), 489–505.

Franceschet, S. and Piscopo, J. M. (2013). Equality, Democracy, and the Broadening and Deepening of Gender Quotas, 9 *Pol. & Gender* 310, 310

Frisbie, W. Parker and Lisa Neidert, 1977, Inequality and the relative size of minority populations: a comparative analysis, American Journal of Sociology 82:1007-30.

Fullenkamp, C., Tenorio, R., Battalio, R., 2003. Assessing individual risk attitudes using field data from lottery games. *Review of Economics and Statistics*.

Gertner, R., 1993. Game Shows and Economic Behavior: Risk-Taking on "Card Sharks." The *Quarterly Journal of Economics* 108, 507–521.

Giles, Michael, 1977, Percent black and racial hostility: an old assumption reexamined. Social Science Quarterly 58:412-17.

Gittler, Joseph, 1956, Understanding Minority Groups. *Wiley*, NY

Goel, A. M., and Thakor, A. V., 2003. Why do firms smooth earnings?
Journal of Business, University of Chicago Press, vol. 76(1), pages 151-192, January.
Gomes, Joao F., 2001, Financing investment, American Economic Review 91, 1263–1288
Gormley, T. A., Gupta, V. K., Matsa, D. A., Mortal, S., and Yang, L., 2020. The Big Three and Board Gender Diversity: The Effectiveness of Shareholder Voice. *European Corporate Governance Institute – Finance Working Paper 714/2020*, Available at ssrn
Gormley, T. A., and Matsa, D. A., 2011. Growing Out of Trouble? Corporate Responses to Liability Risk. *Review of Financial Studies, 2011, 24(8)*
Gormley, T. A., and Matsa, D. A., 2016. Playing it safe? Managerial preferences, risk, and agency Conflicts. *Journal of Financial Economics 122 (3)*
Gormley, T. A., Matsa, D. A., and Milbourn, T. T., 2013. CEO Compensation and Corporate Risk: Evidence from a Natural Experiment. *Journal of Accounting & Economics, 56(2-3)*
Goyal, V.K., and Park, C.W. 2002. Board leadership structure and CEO turnover. *Journal of Corporate Finance* 8, 49–66.
Graham, John R., and Campbell R. Harvey, 2001, The theory and practice of corporate finance: Evidence from the field, Journal of Financial Economics 60, 187— 243.
Gregory-Smith, I., Main, B.G.M., O'Reilly, C.A., 2014. Appointments, pay and performance in UK boardrooms by gender. *Economic Journal* 124.
Harris, Milton, and Artur Raviv, 1991, The theory of capital structure, Journal of Finance 46, 297– 355.
Hartley, Roger, Gauthier Lanot, and Ian Walker, 2005, Who really wants to be a Millionaire? Estimates of risk aversion from game show data, *Working Paper, Department of Economics, University of Warwick.*
Heatherington, E. M., and Feldman, S., 1964. College cheating as a function of subject and situational variables. *J. Educ. Psychol.,* 55 (4), 212-218
Hennessy, Christopher A., and Toni M. Whited, 2005, Debt dynamics, Journal of Finance 60, 1129– 1165.
Hermalin, B.E., 2005. Trends in corporate governance. *Journal of Finance* 60, 2351–2384.
Hermalin, B., Weisbach, M., 1988. The determinants of board composition. RAND Journal of Economics 19, 589–606.
Herminia, I. 2015. Act like a leader, think like a leader. *Harvard Business Press*
Higgs, D., 2003. Review of the role and effectiveness of nonexecutive directors hwww.dti.gov.uk/cld/non_execs_reviewi
Holmstrom, B., 2005. Pay without performance and the managerial power hypothesis: a comment. *Journalof Corporation Law* 30, 703–713.
Homans, George 1974 Social Behavior: Its Elementary Forms, New York: Harcourt, Brace, and Jovanovich

Hosmer, D. W., Lemeshow, S. and Sturdivant, R. X. (2013). Applied Logistic Regression, Third Edition. *Hoboken*, New Jersey: Wiley.

Huang, S., Maharjan, J., and Thakor, A. V., 2020. Disagreement-induced ceo turnover. *Journal of Financial Intermediation*, 43 - 100819

Hughes, P. and Thakor, A. V. 1992. Litigation risk, intermediation, and the underpricing of initial public offerings. *Review of Financial Studies*, 5 – 4, pp. 709 - 742

Jensen, Michael, 1989, The eclipse of the public corporation, Harvard Business Review 67, 61– 75.

Jensen, M.C., Murphy, K.J., 1990. Performance pay and top-management incentives. *Journal of Political Economy* 98, 225–264.

Jianakoplos, N.A., Bernasek, A., 1998. Are women more risk averse? *Economic Inquiry* 36, 620–630.

John, Kose, Lubomir Litov, and Bernard Yeung, 2008, Corporate governance and risk-taking, *The Journal of Finance* 63, 1679–1728.

Jones, J.,1991. Earnings management during import relief investigations. *Journal of Accounting Research* 29, 193–228.

Kang, H., Cheng, M., Gray, S.J., 2007. Corporate governance and board composition: Diversity and independence of Australian boards. Corporate Governance: *An International Review* 15, 194–207.

Kanter, R. M. "Some Effects on Proportions on Group Life: Skewed Sex Ratios and Responses to Token Women." *American Journal of Sociology* 82 (March 1977): 965–90.

Kaplan, S. N., 1994. Campeau's acquisition of Federated: Post-bankruptcy results. *Journal of Financial Economics 35 (1)*

Keldenich, K. and Klemm, M. (2011). Double or Nothing!? Small Groups Making Decisions Under Risk in 'Quiz Taxi' (August 1, 2011). Available at *SSRN*.

Kelley, M. R. and Lemke, R. J. (2013). Decisions under Uncertainty in the Cash Cab *Applied Cognitive Psychology* 27: 542-551

Kim, C. W., and Mauborgne, R. 2015. Blue ocean strategy: How To Create Uncontested Market Space And Make The Competition Irrelevant. *Harvard Business School Publishing*

Kim, I., Pantzalis, C., Park, J.C., 2013. Corporate boards' political ideology diversity and firm performance. *Journal of Empirical Finance* 21, 223–240.

Klein, A., 2002. Audit committee, board of director characteristics, and earnings management. *Journal of Accounting and Economics* 33, 375–400.

Kothari, S.P., Leone, A.J., Wasley, C.E. 2005. Performance matched discretionary accrual measures. *Journal of Accounting and Economics* 39, 163–197.

Kramer, V.W., Konrad, A.M., Erkut, S., and Hooper, M.J., 2007. Critical Mass on Corporate Boards : Why Three or More Women Enhance Governance. *Governance* 0–3.

Lagace, M. , 2001. Can Religion and Business Learn from Each other? *Harvard Business School Working Knowledge* .

Larcker, D.F., Richardson, S.A., Tuna,I. 2007. Corporate governance, accounting outcomes, and organizational performance. *The Accounting Review* 82, 963–1008

Leary, Mark T., and Michael R. Roberts, 2005, Do firms rebalance their capital structures?, *The Journal of Finance* 60, 2575–2619.

Lee, S."Knightian Uncertainty and Capital Structure: Theory and Evidence." Working Paper, George Mason University (2015)

Lehrer, E. , Chiswick, C. , 1993. Religion as a determinant of marital stability. *Demography* 30, 385–404 .

Lehrer, E. L. , 1995. The effects of religion on the labor supply of married women. *Soc. Sci. Res*. 24 (3), 281–301 .

Lehrer, E.L , 2004. Religion as a determinant of economic and demographic behavior in the United States. *Popul. Dev. Rev*. 30 (4), 707–726 .

Li, K., Prabhala, N.R., 2007. Self-selection models in corporate finance. In: Eckbo, B.E.(Ed.), Handbook of Corporate Finance, vol.1: Empirical Corporate Finance. North-Holland, Amsterdam, pp. 37–86.

Linck, J., Netter, J., Yang, T., 2008. The determinants of board structure. Journal of Financial Economics 87, 308–328.

Long, J. S. and Freese, J. (2014). Regression Models for Categorical Dependent Variables Using Stata, Third Edition. College Station, TX: *Stata Press*.

Long, John, and Ileen Malitz, 1985, Investment patterns and Financial Leverage, In Friedman Benjamin, ed.: Corporate Capital Structure in the United States (The University of Chicago Press, Chicago).

Longenecker, J.G. , McKinney, J. , Moore, C. , 2004. Religious intensity, evangelical Christianity, and business ethics: an empirical study. *J. Bus. Ethics* 55 (4), 373–386 .

Marden, Charles and Gladys Meyer, 1973, Minorities in American Society. New York: D. Van Nostrand.

Marsh, Paul, 1982, The choice between equity and debt: An empirical study, Journal of Finance 37, 121– 144

Matsa, D.A., Miller, A.R., 2013. A female style in corporate leadership? Evidence from quotas. *American Economic Journal: Applied Economics* 5, 136–169.

Maxwell, J.C. 2002. Teamwork makes the dream work. *J. Countryman* (Thomas Nilesen Inc.)

Mazar, N. , Amir, O. , Ariely, D. ,2008. The dishonesty of honest people: a theory of self–concept maintenance. *J.Market.Res*.45,633–644 .

McGregor, J., 2020, Goldman Sachs CEO says it won't take a company public without diversity

on McNichols, M.F., 2000. Research design issues in earnings management studies. *Journal of Accounting and Public Policy* 19, 313–345.
its board, *The Washington Post* (WP Company).
Metrick, A., 1995. A natural experiment in `Jeopardy!'. *American economic review* 85, 240–253.
von Meyerinck, Felix, Alexandra Niessen-Ruenzi, Markus Schmid, and Steven Davidoff Solomon, 2018, As California goes, so goes the Nation? The impact of board gender quotas on firm performance and the director labor market, SSRN Electronic Journal.
Michaels, J. W., and Miethe, T., 1989. Applying theories of deviance to academic cheating. *Soc. Sc.,* Q 70, 875-885.
Milbourn, Shockley, and Thakor, 2001. Managerial career concerns and investments in information. *RAND Journal of Economics* Vol. 32, No. 2, Summer 2001 pp. 334–351
Milbourn, T. and Wabara, K., 2021. Religiosity, Higher Purpose, and the Effectiveness of Intense Board Oversight. *Working Paper.*
Miller, A. S. , 2000. Going to hell in Asia: the relationship between risk and religion in a cross–cultural setting. *Rev. Relig. Res.* 40, 5–18 .
Miller, A.S. , Hoffman, J. , 1995. Risk and religion: an explanation of gender differences in religiosity. *J. Sci. Study Relig.* 34, 63–75 .
Monks, R., and Minow, N. 1996, Watching the Watchers (Blackwell Publishers, Cambridge, MA).
Moscovici, S. , Zavalloni, M. , 1969. The group as a polarizer of attitudes. J. Person. *Soc. Psychol.* 12, 125–135 .
Myers, Stewart, 1977, Determinants of corporate borrowing, Journal of Financial Economics 5, 147– 175.
Myers, Stewart C, 1984, The capital structure puzzle, Journal of Finance 39, 575– 592.
Myers, Stewart, 1993, Still searching for optimal capital structure, Journal of Applied Corporate Finance 6, 4–14
Myers, Stewart C, and Nicholas S. Majluf, 1984, Corporate financing and investment decisions when firms have information that investors do not have, Journal of Financial Economics 13, 187– 221.
Noel, D., 1968. A theory of the origin of ethnic stratification. *Social Problems 16*
Norton, E. C., Wang, H., and Ai, C. (2004). Computing interaction effects and standard errors in logit and probit models. *The Stata Journal*, 4(2), pages 154-167.
Osoba, B., 2003. Risk preferences and the practice of religion, evidence from panel data. *Unpublished Working Paper, West Virginia University.*
Ottaviano, G.I.P., Peri, G., 2006. The economic value of cultural diversity: evidence from US cities. Journal of Economic Geography 6, 9–44

Ottaviano, G.I.P., Peri, G., 2008. Immigration and national wages: clarifying the theory and the empirics. Working Paper.

Painter, R., 2021, Why board diversity and the NASDAQ rule requiring it make sense, *CLS Blue Sky Blog*.

Palepu, K. G., 1990. Consequences of leveraged buyouts. *Journal of Financial Economics, 27 (1)*

Pan, Yihui, Stephan Siegel, and Tracy Yue Wang, 2017, Corporate risk culture, *Journal of Financial and Quantitative Analysis* 52, 2327–2367.

Peter, F. and Wagner, A. 2014. The Executive Turnover Risk Premium *Journal of Finance*, vol. 69, no. 4, pp. 1529-1563

Peters, Ryan H., and Lucian A. Taylor, 2017, Intangible capital and the investment-q relation, *Journal of Financial Economics* 123, 251–272.

Post, T., Van Den Assem, M.J., ... Thaler, R.H., 2008. Deal or no deal? decision making under risk in a large-payoff game show. *American Economic Review* 98, 38–71.

Quinn, R. E., and Thakor, A. V. 2019. The Economics of Higher Purpose *Berrett-Koehler Publishers*, ISBN 9781523086405

Quinn, R. E., and Thakor, A. V. 2018. Creating a Purpose-Driven Organization *Harvard Business Review*

Rajan, Raghuram G., and Luigi Zingales, 1995, What do we know about capital structure? Some evidence from international data, *The Journal of Finance* 50, 1421–1460.

Rau, P. R., Sandvik, J., and Vermaelen, T., 2021. Are Women Underpriced? Board Diversity and IPO Performance. *INSEAD Working Paper No. 2021/05/FIN*, Available at ssrn

Rhode, D., Packel, A.K., 2012. Diversity on Corporate Boards: How Much Difference Does Difference Make? *SSRN Electronic Journal*.

Rossi, H., 2014. 7 CEOs with notably devout religious beliefs. *Fortune* 2014/11/11.

Sacerdote, B., and Glaser, E. L., 2008. Education and Religion *Journal of Human Capital*. 2008 vol. 2(2), pages 188-215

Serafeim, G., and Gartenberg, C. 2016. The Type of Purpose That Makes Companies More Profitable. *Harvard Business Review*

Shu, T., Sulaeman, J., and Yeung, P., 2012. Local religious beliefs and mutual fund risk-taking behaviors. *Manag. Sci.*, 58 (10), 1779-1796.

Simmel, Georg (translated by Kurt Wolff),1950, The Sociology of Georg Simmel, New York: Free Press.

Smith, C. P., Ryan, E., and Diggins, D., 1972. Moral decision making: cheating on examinations. *J. Pers.*, 40, 640-660

Song, F., and Thakor, A. V. 2006. Information Control, Career Concerns, and Corporate Governance, *Journal of Fiannce*, https://doi.org/10.1111/j.1540-6261.2006.00891.x

Stark, R., and Finke, R., 2000. Acts of faith: Explaining the Human Side of Religion.

University of California Press .
Stark, O., and Zawojska, E., 2015. Gender differentiation in risk-taking behavior: On the relative risk aversion of single men and single women. *Economics Letters, 137.*
Strebulaev, Ilya A., 2007, Do tests of capital structure theory mean what they say?, *The Journal of Finance* 62, 1747–1787.
Sunden, A. E., Surette, B. J., 1998. Gender differences in the allocation of assets in retirement savings plans. *American Economic Review, 88 (2)*
Terpstra, D. E., Rozell, E. , Robinson, R. , 1993. The influence of personality and demographic variables on ethical decisions related to insider trading. *J. Psychol.* 127 (4), 375–389.
Thakor, A. V. and Quinn, R. E.. 2020. Higher purpose, incentives, and economic performance *ECGI working paper*
Titman, S., 1984. The effect of capital structure on a firm's liquidation decision. *Journal of Financial Economics 13 (1)*
Toy, Norman, Arthur, Stonehill, Lee, Remmers, Richard, Wright, and Theo Beekhuisen, 1974, A comparative international study of growth profitability, and risk as determinants of corporate debt ratios in the manufacturing sector, Journal of Financial and Quantitative Analysis 9, 875– 886.
Tuckman, Bruce W., 1965, Developmental sequence in small groups., *Psychological Bulletin* 63, 384–399.
Tyson, L., 2003. The Tyson report on the recruitment and development of nonexecutive directors. www.london.edu/tysonreport
Vafeas, N.,2005. Audit committees, boards, and the quality of reported earnings. *Contemporary Accounting Research* 22,1093–1122.
Wabara, K., 2005. A fantastic cruise to the zenith: The place of your dreams *Trafford Publishing*
Wabara, K., 2021a. Group Gender Diversity, Individual Power and Influence, and Collective Risk-Taking.*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3804939*
Wang, Y., Clift, B., 2009. Is there a "business case" for board diversity? *Pacific Accounting Review* 21, 88–103.
Warner, J. B., 1977. Bankruptcy Costs: Some Evidence. The Journal of Finance, 32(2), 337–347.
Weber, M., 1905. The Protestant Ethic and the Spirit of Capitalism. *Allen & Unwin,* London.
Weisbach, M.S. 1988. Outside directors and CEO turnover. *Journal of Financial Economics* 20, 431–460.
Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects. *The Stata Journal*, 12(2), pages 308-331.
Williams, K.Y. , O'Reilly, C.A. , 1998. Demography and diversity in organizations: A review of 40 years of research. *Res. Org. Behav.* 20, 77–140.

Yermack, D., 1996. Higher market valuation of companies with a small board of directors. *Journal of Financial Economics* 40, 185–213.

Yetman, N., 1985. Introduction: definitions and perspectives, Pp. 1-20 in N.Yetman (ed.), Majority and Minority: The Dynamics of Race and Ethnicity in American Life. *Allyn and Bacon*, Boston.

Yonker, S. E. (2017). Geography and the market for CEOs. *Manag. Sci.* 63, 609–630

Zahra, S. and Stanton, W., 1998. The Implications of Board of Directors' Composition on Corporate Strategy and Performance, *International Journal of Management*, Vol. 5, No. 2, pp. 229-236.

Zaichkowsky, J. L. (2014). Women in the board room: One can make a difference. International *Journal of Business Governance and Ethics*, 9(1), 91–113.

Zillman, C. (2017). The EU Is Taking a Drastic Step to Put More Women on Corporate Boards. *Fortune*. November 20, 2017.

Zimmer, Lynn, 1988, Tokenism and women in the workplace: The limits of gender-neutral theory, *Social Problems* 35, 64–77.

Zwiebel, Jeffrey, 1996, Dynamic capital structure under managerial entrenchment, American Economic Review 86, 1197– 1215.