# Review of the Literature on Diversity on Corporate Boards

**Jonathan Klick**

APRIL 2021

AMERICAN ENTERPRISE INSTITUTE

# Executive Summary

Despite large increases in the representation of women and people from other minority groups on corporate boards, public and private regulators are pushing for more. California already passed mandates requiring firms headquartered in the state to meet quotas for women and members of other underrepresented groups on their corporate boards. The Nasdaq stock exchange proposed a similar mandate. To support this forced injection of diversity, the regulators point to a wealth of citations claiming diversity improves a firm's value.

Upon examination, though, the research base does not hold up. Many citations come from consulting firm position papers that lack credibility. These reports imply that, because higher-value firms tend to have more-diverse boards, diversity causes the increase in value, without even attempting to adjust for other differences across firms. The academic literature noted in the Nasdaq proposal is not much better. Reliable causal inferences require methods that ensure one is comparing apples to apples, whereas most of the cited literature does little more than add a few control variables to get to an apples-to-bananas comparison, at best.

It also appears that the Nasdaq proposal selectively surveyed the literature on board diversity. When meta-analyses are consulted, the literature as a whole finds little relationship between board diversity and firm value. This systematic review of the literature aligns with numerous other literature reviews, even those performed by individuals predisposed to favor diversity mandates, finding that the evidence is weak for a business case for diversity.

The Nasdaq proposal ignores many studies that are much more reliable methodologically. For example, studies examined the enactment of diversity requirements in Norway, using it as a natural experiment that would provide insight into what happens when firms are forced to diversify their boards. Findings from the Norwegian experience indicate, at best, that diversity mandates do not improve firm value, and some studies find the quotas harmed firm performance. Additionally, many firms chose to go private to avoid the regulation.

There is no credible evidence that diversity requirements systematically improve firm performance.

# Review of the Literature on Diversity on Corporate Boards

## Jonathan Klick

Women are better represented on US corporate boards than ever before. According to data from the proxy advisory firm Institutional Shareholder Services (ISS), by 2019, women held more than one-fourth of board seats in S&P 500 firms and about one-fifth of seats among the Russell 3000.[1] This represents a substantial increase compared to the preceding decade. ISS data also portend even more growth, with women composing almost half of new directors appointed in 2019 in the S&P 500 and the Russell 3000, a more than threefold increase since 2008.[2] In this same period, there was also growth in the inclusion on boards of individuals self-identifying as ethnic minorities.[3]

Despite this organic growth and signs of continued gains in the future, entities such as California[4] and the Nasdaq stock exchange[5] have passed or proposed regulations mandating that firms add women and minority group members to their boards or, in the case of the Nasdaq proposal, provide an explanation for not meeting the requirement.

The California regulation and Nasdaq rule proposal are premised on findings that female board members improve firm performance. For example, California S.B. 826 indicates, "Numerous independent studies have concluded that publicly held companies perform better when women serve on their boards of directors," before summarizing numerous studies from consulting firms such as McKinsey & Company and a few academic studies.[6] Likewise, citing many of the same studies, the Nasdaq proposal asserts, "There is a significant body of research suggesting a positive association between diversity and shareholder value."[7]

As firms are pushed to change their practices to accelerate the already swift trend toward more diversity on boards, it is useful to review the findings of the literature, both those studies cited and more broadly. One concern with the cited literature is its reliance on nonacademic reports from consulting firms that may be influenced by branding considerations and, at minimum, have never been subjected to peer review. With the handful of peer-reviewed studies California and the Nasdaq proposal relied on, it is important to examine whether those studies are representative of the academic literature or cherry-picked.

In this report, I start with a brief, relatively nontechnical primer on empirical work in general, with a focus on causality. As many of the consulting reports admit, their findings cannot answer whether any claimed relationship between firm performance and board makeup represents a causal relationship. While firms with more-diverse boards might perform better than they would with less-diverse boards, the findings could also reflect that more-successful firms might choose diverse board members without the diversity actually affecting performance. Diverse boards might also be concentrated in industries that have happened to do well over the past decade, independently of any contribution by the board. Similarly, almost none of the studies cited explores whether it may be more costly or difficult for some firms to comply with the diversity mandates. A one-size-fits-all approach, as opposed to allowing organic diversity gains, could harm many firms even if the effect on average is positive.

To sort this out, it is necessary to focus on studies that credibly identify the causal impact of board diversity on firm performance. The vast majority of

the studies used to support the diversity regulations do not identify causal effects and, therefore, do not constitute reliable evidence. Among the few studies that provide valid insights into the causal effects of mandating diversity, the evidence is mixed at best. Overall, the literature suggests that such mandates will do little to improve firm performance and may generate losses for shareholders.

## Correlation Is Not Necessarily Causation

The evidence (discussed below) regarding how increasing diversity affects corporate boards either compares an average outcome (e.g., market return) across two or more groups of firms broken down according to the degree of diversity of the firms' boards or uses more-sophisticated techniques (e.g., regression analysis) to compare the relationship between board diversity and outcomes, adjusting for other firm characteristics.

The FCLTGlobal report Nasdaq cited used the general comparison approach.[8] It indicated,

> Looking at MSCI ACWI firms between 2010 and 2017 and using a diversity metric that compasses [*sic*] both age and gender, we found that the most diverse boards (top 20 percent) added 3.3 percentage points to ROIC [return on invested capital], as compared to their least diverse peers (bottom 20 percent).[9]

There are many problems with relying on this approach to support the claim that the Nasdaq proposal will generate firm value. First, there is a relevance problem with taking a claim about a composite diversity index that confounds age and sex and using it as a basis for a regulation focused on sex and minority status. While the FCLTGlobal analysis says gender diversity drives much of the effect (2.6 percent[10]), there is no analysis of minority-status diversity.

Perhaps more importantly, the FCLTGlobal comparison does not account for other potential differences across the companies with the most- and least-diverse boards. For example, during the past decade, the auto industry suffered a slight loss in market return, whereas internet and direct-marketing retail saw growth exceeding 1,000 percent.[11] The primary underlying causes of these diverging prospects obviously have nothing to do with who makes up the company boards in those industries. Car sales are suffering from shifting generational preferences[12] and changing environmental concerns, whereas internet-based retail is a relatively young industry that has benefited from changing technology and other exogenous factors. If car companies have mostly men on their boards and internet-based retail firms have more women on theirs, then a comparison will mechanically generate something like the FCLTGlobal result. At minimum, any such comparison would need to be made on a within-industry basis, to say nothing of needing to adjust for other differences across the firms. Without such adjustments, it is impossible to say anything meaningful about how increasing female participation affects corporate boards.

**To isolate the causal effect of board diversity on outcomes, one needs to compare apples to apples.**

Unfortunately, making the required adjustments is easier said than done. To isolate the causal effect of board diversity on outcomes, one needs to compare apples to apples. Conceptually, the purest test would involve two otherwise identical companies in which one had an all-white, all-male board and the other had a more diverse board. If one were certain the two companies were truly identical, except for their board composition, any differences in outcomes would be either due to random chance or caused by the board differences. Statistically, if the sample size were large enough (many identical firms, some with nondiverse and some with diverse boards), the random component becomes

relatively small on average (and can be bounded), leaving just the board-induced differences.

Of course, it is not possible to examine identical companies since real-life firms differ. If those differences happen to be correlated with board composition (such as the example above positing that boards may differ across industries), determining whether observed outcome differences are due to board composition or these other distinctions will not generally be possible. In statistics, this is called an omitted variable bias. In such a situation, the other differences will confound any estimate of board effects.

To guard against this omitted variable bias, researchers attempt to account for differences across firms. In group-based comparisons, such as the FCLTGlobal example, instead of comparing companies with the most-diverse boards to companies with the least-diverse boards, as suggested, one might make the comparisons in a particular industry. Beyond the industry comparison, one might also try comparing firms with similar corporate governance mechanisms (e.g., comparing firms with staggered boards to other such firms or firms operating under Delaware law with other firms operating this way). This matching or grouping process gets complicated quickly. First, the more attributes an analyst matches on, the smaller the sample gets on which the comparison is performed. Smaller sample sizes increase the influence of random variation in the comparison. In the extreme, it might not be possible to distinguish even large outcome differences among the groups from random noise.

Second, the intuition above is conceptually clear for discrete categories (e.g., firm industry or a particular corporate governance attribute), but grouping by continuous variables is necessary too. If a firm's vintage relates to the outcome metric and the board's diversity, an adjustment is needed. Should the analyst compare only firms that started in the same year? If market capitalization relates to board composition and performance, how close is close enough for grouping purposes? Is the arbitrary distinction between mid- and large-cap firms enough for matching purposes, or should more fine-grained distinctions be made?

These complications must be addressed for FCLTGlobal-type analyses to be taken seriously. Unfortunately, the bulk of these studies, which the Nasdaq proposal relied on, ignore this issue altogether, making their conclusions scientifically unreliable. This leaves the possibility that the studies are vastly overestimating (or underestimating) the causal effect of board diversity on firm outcomes.

Some of the more sophisticated studies use regression techniques to adjust for differences across firms when attempting to isolate how diversity affects corporate outcomes. Generally understood, regression methods "fit" a linear function, relating the control variables (in the current case, the chosen board diversity metric and whatever firm attributes that are to be accounted for) to the outcome variable in an optimal way.[13] For example, if a firm's age is found to have a certain relationship, on average, with the outcome variable and a firm's market capitalization has an estimated relationship with the outcome, then two firms of different ages and market caps can be adjusted to yield an after-adjustment comparison wherein the firms are now conditionally similar, except for their board compositions. After these effects have been accounted for and there are no other differences among the firms, any leftover difference in the firms' outcomes must be due to either random variation or the differing levels of diversity on their boards. Again, if this regression is estimated over many firms, the random component will become relatively less important,[14] and it may be possible to make probabilistic statements about the causal effect of board diversity on the outcome examined.

Although regression techniques are widely used, there is a well-known problem: Regression estimates can be interpreted only as causal effects if the model does not suffer from omitted variable bias. That is, if the model fails to include a control variable (or numerous control variables) that affects the outcome and is correlated with the control variables included in the model, the estimates will not represent the causal effect of a variable on the outcome. For example, even in a regression framework, if one did not account for the industry effects noted above and board diversity differed systematically by industry, the estimated

coefficient on the board diversity variable will include both the true causal effect of board diversity on the outcome and some portion of the industry effects on the outcome. If the regression happens to estimate the true causal effect, it will be entirely accidental, and it is impossible for the researcher to know whether the actual causal effect is bigger, smaller, or the same as the estimated effect is. The researcher cannot even reliably know the sign (i.e., the direction) of the true causal effect.

**Although regression techniques are widely used, there is a well-known problem: Regression estimates can be interpreted only as causal effects if the model does not suffer from omitted variable bias.**

In the example of omitting the industry effects, a simple solution is to adjust by industry. How specifically to define the industry is problematic, as the bias problem could arise if board diversity and outcomes vary at the subindustry—say, four-digit North American Industry Classification System (NAICS)—level, while the researcher adjusts for industry only at the two-digit NAICS level. But conceptually, this is manageable.

The bigger problem is that board diversity and outcomes may be associated with factors the researcher is entirely unaware of, or, sometimes, he or she might be aware of the factor but has no data to make the adjustment. Unobserved heterogeneity is a ubiquitous problem in empirical work, but solving it is crucial to reliably estimating effects.

## When Is Correlation Causation?

Although almost everyone nods in the direction of the causality concerns noted above, many researchers mention it but then progress without taking the implications seriously. For example, the McKinsey report cited in the Nasdaq proposal[15] touts,

> The analysis found a statistically significant relationship between a more diverse leadership team and better financial performance. The companies in the top quartile of gender diversity were 15 percent more likely to have financial returns that were above their national industry median. Companies in the top quartile of racial/ethnic diversity were 35 percent more likely to have financial returns above their national industry median. Companies in the bottom quartile for both gender and ethnicity/race were statistically less likely to achieve above-average financial returns than the average companies in the dataset (that is, they were not just not leading, they were lagging). The results varied by country and industry. Companies with 10 percent higher gender and ethnic/racial diversity on management teams and boards in the US, for instance, had EBIT [earnings before interest and taxes] that was 1.1 percent higher; in the UK, companies with the same diversity level had EBIT that was 5.8 percent higher. Moreover, the unequal performance across companies in the same industry and same country implies that diversity is a competitive differentiator that shifts market share towards more diverse companies.[16]

The report notes, "The relationship between diversity and performance highlighted in the research is a correlation, not a causal link." Almost immediately, the study drops the caution and declares, "More diverse companies are better able to win top talent, and improve their customer orientation, employee satisfaction, and decision making, leading to a virtuous cycle of increasing returns."[17] It could just as

likely be that high-performing firms are better able or more willing to seek out and attract more diverse board members, or there could be some other mediating factor unaccounted for in the analysis that leads to better performance and more diverse boards. Noting that correlation is not causation does not then free one to make such causal claims.

**Noting that correlation is not causation does not then free one to make such causal claims.**

How, then, are causal claims ever possible? Modern statistical and econometric techniques provide some insight.[18] Most modern methods of causal inference take their cue from randomized controlled trials or experiments. If omitted variable bias arises when variables are omitted that influence the outcome (e.g., firm performance) being studied and that are correlated with the "treatment" of interest (e.g., board diversity), one can try to include all the relevant variables. However, failure is guaranteed for the reasons described above. Instead, it makes more sense to ensure somehow that the omitted variables are not correlated with the treatment of interest.

Random assignment of the treatment in an experiment achieves this. If, metaphorically, a coin flip determines whether a firm receives a diverse board, then the existence of a diverse board cannot be associated with firm characteristics such as industry, age, and market cap. Maybe even more importantly, the board assignment will not be correlated with the unquantifiable (but still potentially important) variables such as how forward-looking or progressive a firm is, a firm's risk-taking propensity, and countless other unobservable firm characteristics. If, in this experiment, a firm's board diversity is unrelated to any of the firm's attributes and if one observes firms with more-diverse boards performing better than firms with nondiverse boards are, then the performance differential is either due to random variation (which, as noted earlier, becomes less important as the experiment's sample size grows) or is driven by the presence of the diverse board itself.

This experimental approach is used regularly to test the safety and efficacy of pharmaceuticals and vaccines and in developing other products. In these settings, there are few concerns about whether an observed effect is causal. Unfortunately, it is often impractical to implement this approach for economic policies and regulations. Equal protection constraints and practical considerations limit the extent to which governments can engage in this kind of experimentation in the real world. Lab experiments are sometimes used to examine how mixed-sex groups affect business decision-making,[19] but their artificial settings limit their external validity in extrapolating the results to infer how gender diversity might affect board decision-making in real-world settings with real-world stakes. For example, the stakes involved in the lab experiments are small, and the short duration of the team decision periods may obscure what would happen over longer periods as the decision makers grow more comfortable with each other.

Instead, modern empirical work focuses on quasi-experimental approaches, which mimic the randomization of the lab but in naturalistic settings. Since these are real decision-making settings with real-life stakes, the external validity concerns diminish. In the board diversity literature, there are two main quasi-experimental approaches that have been used to varying degrees of success. For reference when I later describe the studies, I briefly explain their intuition.

The first approach in some of the papers the Nasdaq proposal relied on is a so-called instrumental variables technique.[20] The idea behind instrumental variables is that one finds an instrument (or multiple instruments) that is correlated with the policy variable of interest (in the current case, board diversity) but is otherwise uncorrelated with anything else related to the outcome variable. The first of these conditions allows one to model the policy variable with regression techniques, exploiting that the

instrument is highly related to the policy variable. Because, by assumption, this instrument is otherwise unrelated to the outcome variable, it (as distinct from the policy variable) will not be correlated with any of the unobservable characteristics we worried about above. In practice, the researcher first regresses the policy variable on the instrument (and any other control variables), yielding a model that can be used to predict the policy variable. This prediction of the policy variable is then used in the regression to model the outcome variable (e.g., firm performance). Since the instrument is uncorrelated with the firm's unobservable characteristics, the predicted policy variable (as distinct from the actual policy variable) will likewise be uncorrelated with the unobservable characteristics. Thus, the estimated relationship between the predicted policy variable and the outcome variable will not suffer from omitted variable bias and, therefore, can be interpreted causally.

Although the instrumental variables approach works in theory, practice is a different matter. For starters, the researcher must find a suitable instrument. This instrument needs to strongly correlate with the policy variable being studied, and it must otherwise be unrelated to the outcome variable being examined. The first requirement is testable. Unfortunately, the second criterion is not.[21] At best, the researcher provides an intuitive argument for why he or she believes the instrument is not otherwise related to the outcome variable (except through its effect on the policy variable). If someone can intuit why the instrument is not unrelated to the outcome (or even if he or she cannot, but a reason nonetheless exists), then the instrumental variables analysis should be viewed skeptically. For it to be credible, there should be strong intuitions for both why the instrument is strongly correlated with the policy variable of interest and why the instrument is not otherwise related to the outcome variable being studied. In practice, these intuitions are rarely strong enough to be compelling.

The second approach, not often used in the Nasdaq-cited studies but regularly used in other relevant papers, is more promising. This approach exploits natural experiments. That is, the researcher leverages some outside change in the world that is not initiated (or maybe not even expected) by those affected by it that imposes a change in the policy variable on some firms but not others, as if by random chance. In the current context, the most commonly used natural experiment is the passage of legislation affecting firms in a given jurisdiction as the treatment group, using firms outside the jurisdiction as the control or counterfactual comparison group. Sometimes, these natural experiments might even create within-jurisdiction treatment and control groups through policy exemptions (e.g., a size threshold) or because some firms already inadvertently complied with the rule (e.g., a policy requiring a certain number of women on a board will not affect companies that already have that many women on their boards).

When evaluating these natural experiments, it is important to focus on whether the imposition of the policy shock was random and whether the control group is a suitable counterfactual comparison. Unfortunately, here, too, no diagnostic tests are available to ensure these requirements are satisfied. Intuitively, the more unexpected and less targeted the policy shock is, the more credible the research design and the estimates arising from it are. Likewise, the more comparable the treatment and control groups are, the more confidence one has in the study's findings.

In the finance context, event studies are a common form of a natural experiment.[22] In the standard event study, the event is the policy shock, and the analyst compares how the stock return of a firm (or portfolio of firms) differs relative to what would be expected had the event not occurred. The expectation is estimated using a regression of the firm's returns on various variables (usually including a measure of the overall market return) in the period before the event. This predicted event day (or period) return is netted out of the actual return on the event day, generating the estimate of the event's effect (often called an abnormal or excess return). A similar procedure is used on comparison firms (that are not affected by the event) to rule out the possibility that something other than the event being studied generated the event day effect. In the current context, these event studies could be used to examine the market's reaction to proposed diversity mandates, providing a

"wisdom of crowds"–type estimate of the likely effect of increased diversity on corporate boards.

**Event studies could be used to examine the market's reaction to proposed diversity mandates, providing a "wisdom of crowds"–type estimate of the likely effect of increased diversity on corporate boards.**

While I have simplified this primer for a nontechnical audience, it conveys the main intuitions that can be used to assess empirical analyses of how diverse boards affect firm performance. Conceptually, the closer an analysis is to an apples-to-apples comparison, the more reliable it is. If a study finds that firms having more-diverse boards leads to improved outcomes, then the relevant question is whether diversity actually drives the outcomes in a but-for sense. That is, in the counterfactual world in which the firms did not have diverse boards, would their outcomes be different? Because it is not possible to observe the counterfactual world, it is necessary to rely on comparisons with firms that do not have diverse boards. However, if those nondiverse firms differ along other dimensions, the comparison will not be informative. Omitted variable bias could lead the observed difference to over- or understate the true effect of board diversity on performance. Worse, it is impossible to know even the direction of the true relationship between board diversity and the outcomes being studied, much less the magnitude of the relationship.

To combat this bias, it is tempting to believe that one can adjust for the other differences across firms by either matching firms with diverse boards with similar counterparts whose boards are less diverse or using more-sophisticated regression techniques. However, it is not generally possible to know all the relevant differences. Beyond that, many of the relevant dimensions will not be quantifiable.

In what follows, I examine the studies the Nasdaq proposal relied on and other informative studies ignored in the proposal, categorizing them by how they attempt to address these issues of causal inference. In the first grouping, I look at the studies that make no attempt to address these problems. I do not spend much time on this set since it is wholly unreliable and provides no guidance on how board diversity affects firm outcomes. Next, I cover the studies that attempt to control or adjust for differences across firms. Given the discussion above, these studies are just as unreliable as are those that do nothing to ensure comparability between firms with and without diverse boards. Lastly, I examine the studies that attempt to use some quasi-experimental approach. Because these studies offer the most-credible approaches and, potentially, the most-reliable estimates of the causal effect of board diversity on firm performance, I examine them in detail. With the benefit of this literature review, I then offer general conclusions about the likely effects of the proposed Nasdaq board diversity mandate.

## Studies with Merely Descriptive Comparisons

As discussed above, comparisons of firm outcomes based only on differences in board diversity metrics without adjustments for other differences across firms are not informative. Board diversity might differ coincidentally with many firm attributes that also affect firm outcomes. It is reasonable to suspect that diversity differs by industry, firm age, state of incorporation, and other firm characteristics that also affect

market returns, accounting profits, sales, research and development, and almost everything else one might care to examine as a company outcome. In such a situation, while it may appear that outcomes vary systematically with board diversity, it is just as likely that the relationship is driven by these other variables that are unaccounted for. Any estimated difference will be subject to statistical bias.

As discussed above, the Nasdaq proposal relies on the FCLTGlobal report,[23] which accounts for no differences across firms beyond the diversity of their boards. If board diversity is not randomly distributed across industries (or any other firm characteristic), any difference related to diversity could be driven by these other characteristics. Further, the report provides no test of statistical significance of the reported board diversity effects. This omission is especially notable given that the report mentions a lack of statistical significance with its results regarding firm performance and board member tenure length.[24] It also mentions a lack of a statistically significant relationship between firm returns and whether the firm CEO was a board member.[25]

The MSCI study cited in the Nasdaq proposal[26] indicates that firms with at least three female board members experience gains in return on equity and earnings per share, while firms with no women on their boards saw losses in both metrics.[27] It does not attempt to adjust for any firm characteristics and admits that its small sample size should lead a reader to treat the results with caution.[28] These issues render the study's results wholly unreliable.

The Catalyst report that the Nasdaq proposal relied on proposes that having a sustained large female contingent (three or more board members for at least four of five years) is related to large increases in return on sales, invested capital, and equity.[29] This analysis makes no adjustments across firms, and, rather than examine the overall effect of female board participation, it compares firms with zero women on their boards to firms with three or more women on their boards. As stated earlier, throwing out such variation has little justification and indicates either a lack of statistical sophistication or potentially purposeful data mining.

### Studies with Basic Controls

The Nasdaq proposal cites a 2020 report from the Carlyle Group[30] that indicates that companies in its portfolios with two or more female or minority group directors outperform companies with only one such director, which, in turn, outperform companies with no female or minority directors, controlling for "industry, fund, and vintage year."[31] Unfortunately, the report provides no details on these estimates, nor does it indicate whether these performance differentials are statistically significant. There is no basis to judge these estimates' reliability, even if one were willing to believe that the only variables important to adjust for are industry, fund, and firm age, which would be a largely unjustified assumption.

**The McKinsey authors admit that the relationships observed are correlational, not causal, but they then discuss their results as if the performance metrics they examine can be tied to their board diversity indicators.**

The McKinsey report likewise attempts to use a handful of controls such as firm nationality and a broad-based industry grouping. As noted above, the McKinsey authors admit that the relationships observed are correlational, not causal, but they then

discuss their results as if the performance metrics they examine can be tied to their board diversity indicators. Their analysis has numerous oddities. First, rather than looking at the whole continuous relationship between their diversity indicator and firm performance, they repeatedly simply compare firms in the lowest quartile of diversity with firms in the highest quartile, as if intermediate levels of diversity provide no relevant information. Perhaps intermediate levels of diversity are much more beneficial (or maybe harmful), but it is not possible to know based on the McKinsey analysis.

In an equally odd way, rather than looking at diversity's effect on performance in general, the outcome they study is the likelihood a firm's performance exceeds its nation's industry average. Throwing out variation could obscure important limitations in their findings. For example, if firms with low diversity are trivially below the average and firms with high diversity are trivially above the average, the proper conclusion would be that diversity does not appear to have an effect, even if there is a statistically significant effect on whether a firm is above or below the average. How this analysis is presented makes it is impossible to rule out such a case.[32]

When the McKinsey report does look at a continuous outcome in Exhibit 3, in which EBIT is related to board diversity for the US and Canada, there is no statistically significant relationship between either gender or ethnic diversity of the corporate board and firm performance. This is surprising since the US sample is the largest one examined, which makes it the most reliable of the regions studied. (For the US and Canada, 186 companies were examined, while just 107 UK and 73 Latin American companies were studied.) If there were a robust relationship between board diversity and outcomes, one would have expected to observe it in the US and Canada sample. At best, this suggests that the McKinsey evidence most relevant to US firms does not establish a basis for mandating diverse boards and, at worse, that the statistically significant correlations elsewhere are spurious.

Credit Suisse's gender report cited in the Nasdaq proposal[33] suggests a positive relationship between firm performance and the presence of a woman on a firm's board, adjusting for the firm's sector.[34] The differentials noted in the report (e.g., the 12.2 percent return on equity for firms with at least one female board member vs. the 10.1 percent return for firms with no female directors), however, do not indicate whether they are statistically significant.[35] As with other studies in this section, the Credit Suisse report makes no effort to isolate causality in these relationships.

## Studies That Attempt to Isolate Causation

The Nasdaq proposal frequently cites David A. Carter, Betty J. Simkins, and W. Gary Simpson's study for the proposition that there is a positive relationship between firm value and the presence of women or minorities on a firm's board.[36] Superficially, this study attempts to control for many differences across firms, including differences in total firm size and board size. Of course, since it is never possible to be sure one has made all the necessary adjustments, the authors note that something more is necessary.

> While board diversity could affect firm value, firm value could also affect board diversity. If this is the case, estimation of Equation (1) using OLS [ordinary least squares] can produce biased coefficient estimates. To control for the possibility of endogeneity, we estimate the following system of equations using 2SLS.[37]

Endogeneity is a particular form of the omitted variable bias, and 2SLS is an implementation of the instrumental variables analysis discussed above. However, in implementing 2SLS, the authors do not even attempt to include the necessary instrument. Their Table 4 results illustrate this, as the only variables included in their diversity prediction that are not also included in their firm value equation are the log of the average age of the board (which, by the argument presented in the FCLTGlobal report, directly affects firm value and so is not unrelated to the outcome variable here), an indicator for whether there is a minority board member (in the female

diversity model), and an indicator for whether there is a female board member (in the minority diversity model). Without getting into intuitive arguments about whether an instrument is good, if female board members affect firm value, then a variable capturing that cannot serve as a good instrument for the presence of minority board members and vice versa. That is, even by the logic of the authors' own estimation strategies, their instruments are bad, and therefore their results are not credible.

The Gennaro Bernile, Vineet Bhagwat, and Scott Yonker paper[38] cited in the Nasdaq proposal[39] that indicates board diversity improves many firm outcomes is potentially more credible. It also uses an instrumental variables strategy to account for omitted variable bias. Specifically, the authors use a metric of the diversity of potential directors (defined as people who are serving or have served as directors) who live more than 150 miles from their firms' headquarters but who live near an airport with a nonstop flight to an airport near their headquarters. The intuition is that people agree to be on boards only if it is convenient to participate, which will be a function of transportation ease. If the relevant pool of director candidates who can easily travel to the firm is more diverse, the firm will more successfully attract diverse board members. The data bear this out. The authors find a strong relationship between this pool variable and the diversity of the firms' boards. Through the instrumental variables technique, they show that diversity is associated with many positive firm outcomes.

While this instrument is clever, it does raise concerns. First, even if one assumes arguendo that the empirical strategy is valid, it does not say diverse board members lead to improved outcomes in general. It indicates that diverse candidates who have already served on boards can improve firm outcomes. While this distinction might appear slight, it does make a difference in the context of policies that mandate many firms all chase the existing pool of female and minority board members simultaneously. This research says nothing about how adding female and minority individuals affects a board when such individuals have no previous experience. Second, and more importantly, if firms' outcomes are influenced by factors in their local communities (e.g., agglomeration effects or shared labor markets) and if, all other things being equal, more dynamic and vibrant places attract more transportation linkages because more people want to be there, then the authors' instrument is necessarily capturing effects related to firm outcomes independent of the board member accessibility issue they focus on. If this or anything similar is occurring, then the authors' instrument is no good, and the estimates are not reliable.

In an alternative instrumental variables specification provided in an online appendix,[40] the authors use an instrument that captures average board diversity of a firm's competitors (defined as being a similar size and in the same industry) on the assumption that firms may learn from each other about the benefits of diversity. A first problem with this strategy is that, if a firm's performance is affected by competitors' performance and the existence of a more-diverse board (as is the conclusion of the paper), then, by definition, this instrument is no good. That is, more-diverse boards among competitors both change outcomes in the industry (affecting the firm in that industry) and the firm's likelihood of having a diverse board. Again, the results would be unreliable in this case. A second problem, as discussed above, is that if multiple instruments are available, they could be used simultaneously to allow for calculating the test of overidentifying restrictions, which would provide at least a weak diagnostic of whether the instruments were good. The authors not providing this diagnostic test is a red flag.

If one is skeptical of the authors' instrumental variables strategy but does not wish to throw out the research entirely on this basis, it would be more conservative to examine the authors' regular OLS regression results, which are uniformly much smaller in magnitude, often by a factor of 20 or 30. This suggests a questionable estimation strategy primarily drives the authors' results.

Carter et al.[41] use a fixed effects model to attempt to estimate a causal effect of board diversity on firm performance. A fixed effects model attempts to absorb all fixed unobservable aspects of a firm by including separate baselines for each firm (the so-called fixed effects). This approach works if all relevant

unobservable characteristics are fixed or constant at the firm level. If the unobservable characteristics are changing in a way that is constant across firms in a given period, separate period effects will account for these changes. However, if the unobservable characteristics are changing differentially across firms, the omitted variable bias problem is still present.

**The assumption that current performance is unrelated to previous performance defies belief.**

The authors also attempt a simultaneous equations model (a variant of the instrumental variables technique) to further guard against omitted variable problems. However, as in the earlier Carter, Simkins, and Simpson paper, this approach has problems. Specifically, the authors use lagged outcome variables as their instruments. As with any instrument, if the lagged outcome is related to the current outcome, these instruments will be no good. The assumption that current performance is unrelated to previous performance defies belief. Given the implausibility of the assumptions of Carter et al., their finding that board diversity does not have a statistically significant effect on firm outcomes is not credible.

For similar reasons, Kevin Campbell and Antonio Mínguez-Vera's use of fixed effects models to examine how board diversity affects firm value in a Spanish sample is not credible.[42] As stated before, for the fixed effects model to avoid omitted variable bias, one must assume that either the unobservable heterogeneity across firms is constant or, to the extent it changes, it changes for all firms similarly over time. The authors also attempt an instrumental variables technique, but their instruments are not plausible. For example, one of their instruments is the size of the board of directors. If board size has any effect on firm performance, then their instrumental variables approach does not work. Thus, their mixed conclusions[43] about how women affect boards and other metrics of board diversity are not credible.

One of the better papers cited by the Nasdaq proposal, by Renée B. Adams and Daniel Ferreira, uses both fixed effects and a potentially more plausible instrument in the instrumental variables analysis.[44] The authors instrument the fraction of the firm's board composed of women with a measure of how many female connections male board members have for other boards they sit on. The idea is that knowing more female board members allows women to engage in more networking, leading to an increased likelihood of being on a firm's board. This instrument could be subjected to the concern raised with Bernile, Bhagwat, and Yonker's secondary instrument. Namely, if a competitor's performance affects the firm's performance and if more women on a board affect the competitor's performance, then the instrument would not be unrelated to the outcome being studied. However, Adams and Ferreira do not restrict attention to connections to women made through competitors' boards, so any concern of this type might be mitigated. This represents a reasonable strategy. While Adams and Ferreira show that increasing female board membership improves board attendance by all board members and improves other monitoring metrics, the ultimate effect on firm value appears detrimental, sometimes to a statistically significant degree.

Bin Srinidhi, Ferdinand A. Gul, and Judy Tsui use an instrumental variables technique (specifically a Heckman selection model) to examine how female directors affect the transparency or quality of a firm's earnings data, focusing on accruals estimation errors by the firm and indicators of manipulation or excessive management of earnings announcements.[45] As cited in the Nasdaq proposal, they find that female participation on a firm's board improves the indicators of earnings data quality.[46] As with much of this literature, the authors do little to discuss or justify why their identification strategy supposedly works. Most of the variables in the first stage of the analysis are explicitly related to firm performance and

attributes.[47] Clearly, these variables can directly affect the outcome variable and therefore cannot serve as good instruments for identification purposes. The only plausible candidate is the inclusion of the percentage of women employed in the firm's industry. However, since women are not randomly distributed across industries and since firms in an industry likely mimic each other in many things related to earnings, earnings management, and earnings reporting (e.g., using the same outside auditor[48]), this candidate instrument also is likely related to the outcome variables examined in the paper. As I have repeatedly noted, this concern undercuts the reliability of the paper's empirical conclusions.

**Of necessity, matching can be carried out using only observable characteristics, since it is impossible to know whether the firms are similar on unobserved dimensions.**

María Consuelo Pucheta-Martínez, Immaculada Bel-Oms, and Gustau Olcina-Sempere's paper also looks at transparency metrics (specifically, measures of audit report quality).[49] Although the authors find that their measures of female board participation are associated with better audit quality metrics in Spanish data, they make no attempt to account for unobservable characteristics, merely controlling for observable firm characteristics. This leaves no confidence that their results represent causal effects.

Similarly, the Nasdaq proposal cites[50] Francisco Bravo and Maria Dolores Alcaide-Ruiz's finding that, although female participation on a firm's audit committee does not affect a firm's propensity to disclose forward-looking financial information, having women with financial expertise on the committee improves this propensity.[51] Once again, however, this analysis does nothing to account for omitted variable bias and lacks credibility.

Lawrence J. Abbott, Susan Parker, and Theresa J. Presley find that firms with at least one female director are less likely to issue financial restatements than firms with no women on their board are.[52] Their attempt to isolate causality involves matching each firm with a female board member with a comparable firm from the same industry (similar size, type of audit firm used, etc.) with no women on the board. Matching approaches such as this are similar to regression techniques but allow for a type of nonlinear modeling of the effect of the match or control variables. However, of necessity, matching can be carried out using only observable characteristics, since it is impossible to know whether the firms are similar on unobserved dimensions. Thus, generally, matching does not address omitted variable bias.

Aida Sijamic Wahid's article[53] also attempts to examine the transparency of firms with female representation on their boards by examining financial reporting mistakes and fraud indicators. The article uses instrumental variables techniques to isolate causality and finds that firms with women on their boards engage in less fraud and make fewer reporting mistakes. In addition to the instrumental variables approach, the article uses fixed effects models, though it never combines the fixed effects and instrumental variables approach, which would be the most rigorous approach. Wahid's instruments for female participation are the female population around the firm's headquarters and the longitude measurement at the firm's headquarters. While the first instrument has an intuitive explanation—namely, firms located where there are more women may find it easier to solicit female board members—the longitude instrument is not intuitive and smacks of data mining (i.e., opportunistically searching for an instrument that provides particular results). Because neither instrument is likely to vary much

(or at all for longitude) at the firm level from year to year, it becomes obvious why Wahid does not estimate the instrumental variables regressions with fixed effects. That is, with no variation, estimation becomes impossible. Because of this, however, any firm characteristics related to a firm's location (e.g., more-talented CEOs may prefer to live in certain locations) will cause the instruments to fail, leaving the estimates without credibility.

Using Chinese firms, Douglas Cumming, T. Y. Leung, and Oliver Rui examine how gender diversity affects the likelihood that a firm engages in fraud.[54] They find that firms with a higher fraction of women on the board engage in less fraud, although the effects are smaller (and sometimes not statistically significant) in female-dominated industries. In an attempt to determine causality, their paper uses an instrumental variables technique but proceeds to use firm characteristics, including characteristics of the firm's chairperson and general manager, board, and ownership structure. Obviously, all these characteristics can directly affect the likelihood of fraud (e.g., one of the characteristics used is frequency of board meetings, which should help monitor fraud) and so provide no confidence in a causal interpretation. The paper also discusses that one would find similar results with other forms of diversity,[55] but since the authors do not study this, it is pure speculation.

Gul, Srinidhi, and Anthony C. Ng's 2011 paper also examines the informativeness of firms with greater female representation on boards.[56] It concludes that firms with more female representation on their boards have more informative stock prices, as measured by idiosyncratic volatility and "future earnings incremental explanatory power."[57] While this paper uses longitudinal data, it does not estimate fixed effects models for its regressions even though that would account for more unobservable characteristics of the firms studied. In addition to this bias point, because fixed effects absorb much of the variation in the data, many of the paper's borderline statistically significant effects (if not the others) would likely no longer be statistically significant (e.g., regressions 9, 10, and 12 in Table 4). It is puzzling why the authors did not examine the more standard (and credible) fixed effects regression specification given there appear to be no data limitations in doing so.

The paper does exploit a potentially interesting natural experiment involving Norwegian legislation that required firms to have more female directors. After examining how many additional female directors 75 Norwegian firms added between 2005 and 2009, the authors report that idiosyncratic volatility increased as more female directors were added and the effect was statistically significant. Once again, the authors choose not to examine the fixed effects model, which would be more credible. Further, they artificially focus only on Norwegian firms, when they could have easily used non-Norwegian firms in a more standard natural experiment framework wherein non-Norwegian firms serve as the comparison or control group. Such an analysis would be more informative and reliable.

Although ignored in the Nasdaq proposal, a 2019 paper by Philip Yang et al. offers a more thorough examination of the Norwegian experience.[58] It uses the natural experiment in Norway to compare the performance of Norwegian firms with a control group of firms in other Scandinavian countries. Paying close attention to whether the conditions for a natural experiment are met, the authors find that the regulation requiring more women on Norwegian boards led to worse firm performance and greater firm risk. These results are consistent with the findings of Kenneth R. Ahern and Amy K. Dittmar, who found that firms that were more affected by the Norwegian regulation (i.e., had to appoint more women) suffered statistically significant declines in stock price, firm value, and other measures of operating performance and exhibited greater risk after the adoption of the mandate.[59]

David A. Matsa and Amalia R. Miller provide additional support,[60] comparing Norwegian firms to firms in other Scandinavian countries and private Norwegian companies that were not subject to the rule. This variation allows the authors to rule out any possibility that a nonregulation-related event in Norway may have been affecting firms subjected to the regulation. Across many specifications, Matsa and Miller find robust evidence that the regulation led to declining profits and the effect was statistically significant. This

decline was not observed among the unaffected Norwegian private firms nor among the unaffected firms from other countries.

Given this evidence of the bad effects of the Norwegian mandate, it is not surprising that Øyvind Bøhren and Siv Staubo found that about half of Norwegian firms that were going to be affected by the mandate changed their status (i.e., went private) to avoid the regulation.[61] In a more recent working paper, B. Espen Eckbo, Knut Nygaard, and Karin S. Thorburn make different modeling choices (e.g., looking at different time windows and examining the effect of heterogeneity) and find results that are more aligned with the conclusion that the Norwegian mandate did not affect stock returns or other measures of firm value in a statistically significant way.[62]

Taking these findings on transparency and informativeness, David Abad et al. examine information asymmetries via bid-ask spreads, the idea being that, if firms are more transparent, there is less concern of trading by relatively better-informed insiders.[63] When outsiders worry about information deficits, markets become less liquid, leading to larger bid-ask spreads. Analyzing Spanish data, the authors find that firms with better female representation on boards exhibit smaller bid-ask spreads, again suggesting that such firms are more transparent. The authors use a generalized method of moments (GMM) estimation technique to account for omitted variable bias. GMM is a variant of instrumental variables and requires good instruments. As is common in GMM approaches, the authors use lagged or differenced versions of their model's variables as instruments and (distinct from the other papers reviewed here) appropriately report the diagnostic test of overidentifying restrictions, which indicates the instruments are good.

However, as discussed in the empirical primer, this diagnostic test "works" only if one believes the various instruments are not subject to similar omitted variable effects. For example, if one believes the model's variables lags (i.e., the observation from the prior period) are subject to similar random shocks (or something related), the instruments will pass the diagnostic test but still not work as instruments that surmount the omitted variable bias. If there is unexplained persistence in the instruments that could be related to unobservable characteristics that also affect the outcome variable, the GMM approach does not solve the bias concern.

**The Nasdaq proposal clearly cherry-picks studies based on whether they claim to find positive outcomes associated with an increase in female representation on boards and a smattering of articles looking at other kinds of diversity.**

The same concern undercuts the conclusions of Maria Encarnación Lucas-Pérez et al., who relate female participation on boards with better controls on executive pay using a similar GMM approach with lagged variables as the instruments.[64] Unless one is willing to assume a firm's past characteristics are unrelated to the outcome variables, GMM's use of lagged variables as instruments does not overcome the omitted variable bias problem.

### Meta-Analyses

The studies cited in the Nasdaq proposal are clearly not an exhaustive review of the literature, nor are they a random sampling. The Nasdaq proposal clearly

cherry-picks studies based on whether they claim to find positive outcomes associated with an increase in female representation on boards and a smattering of articles looking at other kinds of diversity. As discussed above, the proposal also cannot claim to focus on methodologically sound studies, given the generally poor quality of the studies invoked.

Many other studies in the literature find no effect, or even a negative effect, from increased board diversity. For example, Kathleen A. Farrell and Philip L. Hersch find a negative, though statistically insignificant, effect of appointing a new female board member on firms' stock returns.[65] Interestingly, this paper also clearly demonstrates that adding a female board member is well predicted by firm outcomes. That is, better-performing firms are more likely to add a woman when a new board seat becomes available. This highlights the omitted variable concern raised above, suggesting that any study failing to account for the endogeneity of female board representation is destined to yield noncredible results.

In another example, Caspar Rose finds a negative (though statistically insignificant) relationship between female board representation and firm value in Danish data.[66] However, this is not done well because it, too, does not account for omitted variable bias. The methods are not substantially different from those used in many of the Nasdaq-cited studies described above. Even using Danish data cannot be cited as the reason for exclusion, given Nasdaq's repeated use of studies with international data (e.g., multiple studies cited use Spanish data).

One is left with the distinct impression that the proposal's statement "**Nasdaq reviewed dozens of empirical studies and found that an extensive body of academic research demonstrates that diverse boards are positively associated with improved corporate governance and financial performance**"[67] (emphasis in original) is not altogether accurate, since clearly Nasdaq ignored evidence that did not support (or even contradicted) its proposal. The review of the evidence on how women affect corporate boards by noted feminist scholar Deborah Rhode and Amanda K. Packel provides an interesting contrast: "After exploring the strengths and limitations of various methodological approaches and survey findings, [we conclude] that the relationship between diversity and financial performance has not been convincingly established."[68]

For a more comprehensive, aggregate view of the literature, one can examine what meta-analyses of the articles find. Meta-analyses attempt to define the literature according to certain criteria and then use the empirical findings as a dataset, providing average results across papers and sometimes providing what are essentially regression analyses of the other papers' regression results to see what factors appear to influence the findings in general. Meta-analyses are not without flaws, and there are reasonable grounds for criticisms. Choosing what articles constitute a literature can be arbitrary, though the better meta-analyses lay out specific criteria and document how they searched for articles that fit. A more substantive concern involves treating studies of widely differing methodological quality comparably (with each study counting as an equal data point). However, meta-analyses can provide a useful and somewhat more objective summary of a literature. Meta-analyses can also sometimes be used to identify publication biases, such as the tendency of journals to accept only articles with statistically significant or ideologically attractive results.

The Nasdaq proposal cites a few meta-analyses but does not appear to internalize the message that many of them suggest a different story from its selective review of the literature. For example, an analysis by Jan Luca Pletzer et al. of 20 studies finds that the average effect estimated by the studies regarding female participation on boards and firm performance is small and statistically insignificant.[69] Corrine Post and Kris Byron's 2015 meta-analysis of 140 studies (45 using US data) found that, on average, there was a small, positive, and highly variable (perhaps providing evidence of model mis-specification concerns) relationship between female board participation and accounting measure-based returns. But there was no relationship, on average, between female board members and market returns.[70] The authors also present a meta-analysis of the relationship between women on boards and the likelihood a firm engages in stakeholder facing, or

so-called socially responsible business practices, finding a positive relationship on average.[71]

The mixed relationship, at best, between female board participation and firm outcomes that results from a less selective analysis of the literature contrasts with the overwhelmingly positive picture painted in the Nasdaq proposal. Alice H. Eagly chastises the ideologically motivated cherry-picking that seems to infect documents such as the Nasdaq proposal, writing, "Despite advocates' insistence that women on boards enhance corporate performance and that diversity of task groups enhances their performance, research findings are mixed, and repeated meta-analyses have yielded average correlational findings that are null or extremely small."[72]

## Wishful Thinking

A more complete and nuanced view of the literature on female participation on corporate boards suggests that the literature provides no strong evidentiary basis for requiring firms to increase female participation. The papers purporting to find large gains to increasing gender diversity suffer from crippling methodological flaws. It is clear from the data that firms do not randomly decide to appoint women. Many performance-related characteristics predict whether a firm fills an open board seat with a woman. This requires an empirical researcher to take the omitted variable bias question seriously. Many of the attempts in the literature to use instrumental variables techniques to overcome selection and endogeneity problems are not well-thought-out. In some cases, the papers try to implement instrumental variables without using any instruments; in other cases, the papers use instruments that their own analyses suggest directly affect the outcome variable being studied. Therefore, they are invalid instruments.

Perhaps the better way forward is examining the handful of natural experiments that exist, such as the Norwegian experience. As discussed, the papers that do the most–methodologically sophisticated analyses of this experience, such as by Ahern and Dittmar and Matsa and Miller, demonstrate that the mandate worsened firm performance and value and that the effects are statistically significant and economically large. At best, Eckbo and Nygaard's results indicate the Norwegian experiment had no systematic effect on firm performance. Further, Bøhren and Staubo's analysis indicates firms are willing to make organizational changes rather than be forced to alter their boards. If this experience can be generalized to the United States, one might predict that mandates such as the Nasdaq proposal or California's board diversity regulations will lead firms to switch exchanges or the location of their headquarters or go private. Organic growth in the numbers of women and people from other underrepresented groups likely is the best approach to achieving more-diverse corporate boards.

## About the Author

**Jonathan Klick** is a professor of law at the University of Pennsylvania and the Erasmus Chair of Empirical Legal Studies at the Erasmus University Rotterdam. He has held visiting positions at Yale Law School, Columbia Law School, Northwestern Pritzker School of Law, and the University of Southern California Gould School of Law. He was the Jeffrey A. Stoops Professor of Law and Economics at Florida State University.

# Notes

1. Kosmas Papadopoulos, "U.S. Board Diversity Trends in 2019," Institutional Shareholder Services, ISS Analytics, May 31, 2019, 4, https://www.issgovernance.com/file/publications/ISS_US-Board-Diversity-Trends-2019.pdf.
2. Papadopoulos, "U.S. Board Diversity Trends in 2019," 3.
3. Papadopoulos, "U.S. Board Diversity Trends in 2019," 4.
4 S.B. 826, 2017–18, Reg. Sess. (CA 2018), https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB826; and A.B. 979, 2019–20, Reg. Sess. (CA 2020), https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB979.
5. Nasdaq Stock Market LLC, "Proposed Rule Change to Adopt Listing Rules Related to Board Diversity" (Form 19b-4), December 1, 2020, https://listingcenter.nasdaq.com/assets/RuleBook/Nasdaq/filings/SR-NASDAQ-2020-081.pdf.
6. S.B. 826, 2017–18, Reg. Sess. (CA 2018), § 1.
7. See Nasdaq Stock Market LLC, "Proposed Rule Change to Adopt Listing Rules Related to Board Diversity," 16.
8. Nasdaq Stock Market LLC, "Proposed Rule Change to Adopt Listing Rules Related to Board Diversity," 17.
9. FCLTGlobal, *The Long-Term Habits of a Highly Effective Corporate Board*, March 29, 2019, 11, https://www.fcltglobal.org/resource/the-long-term-habits-of-a-highly-effective-corporate-board/.
10. FCLTGlobal, *The Long-Term Habits of a Highly Effective Corporate Board*, 12.
11. Fidelity Investments, "Sectors & Industries—Performance," https://eresearch.fidelity.com/eresearch/markets_sectors/sectors/si_performance.jhtml?tab=siperformance.
12. See, for example, Adrienne Roberts, "Driving? The Kids Are So Over It," *Wall Street Journal*, April 20, 2019, https://www.wsj.com/articles/driving-the-kids-are-so-over-it-11555732810.
13. In the standard case of ordinary least squares regression, the parameters or coefficients for the control variables are chosen to minimize the sum of the squared errors, in which the errors are calculated as the difference between the regression's prediction and the actual outcome for each observation. By squaring the errors, the regression treats positive and negative errors of the same magnitude equivalently.
14. Under reasonable assumptions, it will be possible to engage in hypothesis testing in which a researcher can ask what the likelihood is that he or she would observe the estimated relationship between board diversity and the studied outcome if, in reality, the relationship is zero. If this likelihood is low (conventionally, less than 5 percent), the researcher will conclude that the estimated difference is statistically significant; that is, the difference is too large to be explained by random variation alone. Alternatively, the researcher might provide an estimate of how unlikely it is that one would observe an estimated effect as large or larger in magnitude than the study's finding is, if the "true" effect were zero. This estimate of unlikeliness is referred to as a *p* value. Many of the results relied on in the Nasdaq proposal do not provide a *p* value or even statements about the statistical significance of their findings. These omissions make it impossible to know whether the reported outcome differences between diverse and nondiverse firms are anything but random noise.
15. Nasdaq Stock Market LLC, "Proposed Rule Change to Adopt Listing Rules Related to Board Diversity," 17.
16. Vivian Hunt, Dennis Layton, and Sara Prince, *Diversity Matters*, McKinsey & Company, February 2, 2015, 1, https://www.mckinsey.com/~/media/mckinsey/business%20functions/organization/our%20insights/why%20diversity%20matters/diversity%20matters.ashx.
17. Hunt, Layton, and Prince, *Diversity Matters*, 1.
18. For a more technical discussion of the details in this section, see Jonah B. Gelbach and Jonathan Klick, "Empirical Law and Economics," in *The Oxford Handbook of Law and Economics, Volume I: Methodology and Concepts*, ed. Francesco Parisi (Oxford, UK: Oxford University Press, 2017).
19. See, for example, Vicki L. Bogan, David R. Just, and Chekitan S. Dev, "Team Gender Diversity and Investment Decision Making Behavior," *Review of Behavioral Finance* 5, no. 2 (February 2013): 134–52, https://www.researchgate.net/publication/228204293_Team_Gender_Diversity_and_Investment_Decision_Making_Behavior. The authors show that increasing gender diversity in teams has

mixed effects on risk-taking behavior.

20. For a technical discussion of instrumental variables, see Jeffrey M. Wooldridge, *Introductory Econometrics: A Modern Approach*, 4th ed. (Mason, OH: South-Western Cengage Learning, 2009), chap. 15.

21. In principle, if one can identify many instruments, a test for overidentifying restrictions exists. The intuition for the test assumes that at least one of the instruments satisfies both conditions (related to the policy variable and otherwise unrelated to the outcome variable). The test examines the estimates from using each instrument separately and in each combination and then examines whether all the estimates are statistically equivalent. The idea is that, if one has found the "right" answer and all the instruments are good instruments, each should lead to the same outcome. If different outcomes are estimated, then at least one of the instruments (and maybe all) is bad. The problem, however, is that, if all the instruments are similarly correlated with the unobservable characteristics (a condition that cannot be tested since it is not possible to examine a correlation with an unobserved quantity), they will all lead to the same biased estimate, passing the test without validating the estimates.

22. For a relatively thorough but accessible overview of event studies, see Jill E. Fisch, Jonah B. Gelbach, and Jonathan Klick, "The Logic and Limits of Event Studies in Securities Fraud Litigation," *Texas Law Review* 96, no. 3 (February 2018): 553–618, https://texaslawreview.org/logic-limits-event-studies-securities-fraud-litigation/.

23. FCLTGlobal, *The Long-Term Habits of a Highly Effective Corporate Board*, 11.

24. FCLTGlobal, *The Long-Term Habits of a Highly Effective Corporate Board*, 12.

25. FCLTGlobal, *The Long-Term Habits of a Highly Effective Corporate Board*, 10.

26. Nasdaq Stock Market LLC, "Proposed Rule Change to Adopt Listing Rules Related to Board Diversity," 18.

27. Meggin Thwing Eastman, Damion Rallis, and Gaia Mazzucchelli, *The Tipping Point: Women on Boards and Financial Performance*, MSCI, December 2016, https://www.msci.com/documents/%2010199/fd1f8228-cc07-4789-acee-3f9ed97ee8bb.

28. Thwing Eastman, Rallis, and Mazzucchelli, *The Tipping Point*, 7.

29. Harvey M. Wagner and Nancy M. Carter, "The Bottom Line: Corporate Performance and Women's Representation on Boards (2004–2008)," Catalyst, March 1, 2011, https://www.catalyst.org/research/the-bottom-line-corporate-performance-and-womens-representation-on-boards-2004-2008/.

30. Nasdaq Stock Market LLC, "Proposed Rule Change to Adopt Listing Rules Related to Board Diversity," 16, 144.

31. Jason M. Thomas and Megan Starr, *From Impact Investing to Investing for Impact*, Carlyle Group, February 24, 2020, 5, https://www.carlyle.com/global-insights/invest-with-impact-whitepaper-thomas-starr.

32. Hunt, Layton, and Prince, *Diversity Matters*, Exhibits 1 and 2. Similar criticisms apply to the updated analyses presented in Sundiatu Dixon-Fyle et al., *Diversity Wins: How Inclusion Matters*, McKinsey & Company, May 19, 2020, 13, https://www.mckinsey.com/featured-insights/diversity-and-inclusion/diversity-wins-how-inclusion-matters.

33. See, for example, Nasdaq Stock Market LLC, "Proposed Rule Change to Adopt Listing Rules Related to Board Diversity," 18.

34. Julia Dawson, Richard Kersley, and Stefano Natella, *The CS Gender 3000: Women in Senior Management*, Credit Suisse, September 2014.

35. Dawson, Kersley, and Natella, *The CS Gender 3000*, 16.

36. See, for example, Nasdaq Stock Market LLC, "Proposed Rule Change to Adopt Listing Rules Related to Board Diversity," 17.

37. David A. Carter, Betty J. Simkins, and W. Gary Simpson, "Corporate Governance, Board Diversity, and Firm Value," *Financial Review* 38, no. 1 (February 2003): 43, https://onlinelibrary.wiley.com/doi/10.1111/1540-6288.00034.

38. Note the paper cited is the working paper, which has since been published as Gennaro Bernile, Vineet Bhagwat, and Scott Yonker, "Board Diversity, Firm Risk, and Corporate Policies," *Journal of Financial Economics* 127, no. 3 (March 2018): 588–612, https://www.sciencedirect.com/science/article/abs/pii/S0304405X17303215.

39. See Nasdaq Stock Market LLC, "Proposed Rule Change to Adopt Listing Rules Related to Board Diversity," 17–18.

40. Gennaro Bernile, Vineet Bhagwat, and Scott Yonker, "Internet Appendix to 'Board Diversity, Firm Risk, and Corporate Policies,'" *Journal of Financial Economics* 127, no. 3 (March 2018), http://jfe.rochester.edu/Bernile_Bhagwat_Yonker_app.pdf.

41. David A. Carter et al., "The Gender and Ethnic Diversity of US Boards and Board Committees and Firm Financial Performance," *Corporate Governance: An International Review* 18, no. 5 (September 2010): 396–414, https://onlinelibrary.wiley.com/doi/abs/

10.1111/j.1467-8683.2010.00809.x.

42. Kevin Campbell and Antonio Mínguez-Vera, "Gender Diversity in the Boardroom and Firm Financial Performance," *Journal of Business Ethics* 83, no. 3 (December 2008): 435–51, https://link.springer.com/article/10.1007/s10551-007-9630-y.

43. The Nasdaq proposal cites the study's conclusion that "at a minimum . . . increased gender diversity can be achieved without destroying shareholder value," but given the methodological flaws, even this conclusion is too strong. Nasdaq Stock Market LLC, "Proposed Rule Change to Adopt Listing Rules Related to Board Diversity," 148.

44. Renée B. Adams and Daniel Ferreira, "Women in the Boardroom and Their Impact on Governance and Performance," *Journal of Financial Economics* 94, no. 2 (November 2009): 291–309, https://personal.lse.ac.uk/ferreird/gender.pdf.

45. Bin Srinidhi, Ferdinand A. Gul, and Judy Tsui, "Female Directors and Earnings Quality," *Contemporary Accounting Research* 28, no. 5 (December 2011): 1610–44, https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1911-3846.2011.01071.x.

46. Nasdaq Stock Market LLC, "Proposed Rule Change to Adopt Listing Rules Related to Board Diversity," 151.

47. See Srinidhi, Gul, and Tsui, "Female Directors and Earnings Quality," 1617.

48. See, for example, Ahsan Habib, "Audit Firm Industry Specialization and Audit Outcomes: Insights from Academic Literature," *Research in Accounting Regulation* 23, no. 2 (October 2011): 114–29, https://www.sciencedirect.com/science/article/abs/pii/S1052045711000373.

49. María Consuelo Pucheta-Martínez, Inmaculada Bel-Oms, and Gustau Olcina-Sempere, "Corporate Governance, Female Directors and Quality of Financial Information," *Business Ethics: A European Review* 25, no. 4 (October 2016): 363–85, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2836230.

50. Among other places, see Nasdaq Stock Market LLC, "Proposed Rule Change to Adopt Listing Rules Related to Board Diversity," 25.

51. Francisco Bravo and Maria Dolores Alcaide-Ruiz, "The Disclosure of Financial Forward-Looking Information: Does the Financial Expertise of Female Directors Make a Difference?," *Gender in Management* 34, no. 2 (March 2019): 140–56, https://www.researchgate.net/publication/331744002_The_disclosure_of_financial_forward-looking_information_Does_the_financial_expertise_of_female_directors_make_a_difference.

52. Lawrence J. Abbott, Susan Parker, and Theresa J. Presley, "Female Board Presence and the Likelihood of Financial Restatement," *Accounting Horizons* 26, no. 4 (December 2012): 607–29, https://www.researchgate.net/publication/274429063_Female_Board_Presence_and_the_Likelihood_of_Financial_Restatement.

53. Aida Sijamic Wahid, "The Effects and the Mechanisms of Board Gender Diversity: Evidence from Financial Manipulation," *Journal of Business Ethics* 159, no. 3 (October 2019): 705–25, https://link.springer.com/article/10.1007/s10551-018-3785-6.

54. Douglas Cumming, T. Y. Leung, and Oliver Rui, "Gender Diversity and Securities Fraud," *Academy of Management Journal* 58, no. 5 (October 2015): 1572–93, https://journals.aom.org/doi/abs/10.5465/amj.2013.0750.

55. Nasdaq Stock Market LLC, "Proposed Rule Change to Adopt Listing Rules Related to Board Diversity," 154.

56. Ferdinand A. Gul, Bin Srinidhi, and Anthony C. Ng, "Does Board Gender Diversity Improve the Informativeness of Stock Prices?," *Journal of Accounting and Economics* 51, no. 3 (April 2011): 314–38, https://www.sciencedirect.com/science/article/abs/pii/S0165410111000176.

57. Essentially, this measures how much of an improvement one gets in the R squared of a firm's returns when future earnings terms are added to the regression. Gul, Srinidhi, and Ng, "Does Board Gender Diversity Improve the Informativeness of Stock Prices?"

58. Philip Yang et al., "Women Directors, Firm Performance, and Firm Risk: A Causal Perspective," *Leadership Quarterly* 30, no. 5 (October 2019), https://www.sciencedirect.com/science/article/pii/S1048984318303217.

59. Kenneth R. Ahern and Amy K. Dittmar, "The Changing of the Boards: The Impact on Firm Valuation of Mandated Female Board Representation," *Quarterly Journal of Economics* 127, no. 1 (February 2012): 137–97, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1364470.

60. David A. Matsa and Amalia R. Miller, "A Female Style in Corporate Leadership? Evidence from Quotas," *American Economic Journal: Applied Economics* 5, no. 3 (July 2013): 136–69, https://www.aeaweb.org/articles?id=10.1257/app.5.3.136. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1636047.

61. Øyvind Bøhren and Siv Staubo, "Does Mandatory Gender Balance Work? Changing Organizational Form to Avoid Board Upheaval," *Journal of Corporate Finance* 28 (October 2014): 152–68, https://www.sciencedirect.com/science/article/abs/pii/S0929119913001302.

62. B. Espen Eckbo, Knut Nygaard, and Karin S. Thorburn, "Valuation Effects of Norway's Board Gender-Quota Law Revisited," *Management Science* (forthcoming), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2746786.

63. David Abad et al., "Does Gender Diversity on Corporate Boards Reduce Information Asymmetry in Equity Markets?," *BRQ Business Research Quarterly* 20, no. 3 (July–September 2017): 192–205, https://www.sciencedirect.com/science/article/pii/S2340943617300142.

64. Maria Encarnación Lucas-Pérez et al., "Women on the Board and Managers' Pay: Evidence from Spain," *Journal of Business Ethics* 129, no. 2 (June 2015): 265–80, https://link.springer.com/article/10.1007/s10551-014-2148-1.

65. Kathleen A. Farrell and Philip L. Hersch, "Additions to Corporate Boards: The Effect of Gender," *Journal of Corporate Finance* 11, nos. 1–2 (March 2005): 85–106, https://www.sciencedirect.com/science/article/abs/pii/S0929119904000264.

66. Caspar Rose, "Does Female Board Representation Influence Firm Performance? The Danish Evidence," *Corporate Governance: An International Review* 15, no. 2 (March 2007): 404–13, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=972533.

67. Nasdaq Stock Market LLC, "Proposed Rule Change to Adopt Listing Rules Related to Board Diversity," 9.

68. Deborah L. Rhode and Amanda K. Packel, "Diversity on Corporate Boards: How Much Difference Does Difference Make?," *Delaware Journal of Corporate Law* 39, no. 2 (2014): 377–426, https://law.stanford.edu/publications/diversity-on-corporate-boards-how-much-difference-does-difference-make-2/.

69. Jan Luca Pletzer et al., "Does Gender Matter? Female Representation on Corporate Boards and Firm Financial Performance—A Meta-Analysis," *PLOS ONE* 10, no. 6 (2015), https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0130005.

70. Corinne Post and Kris Byron, "Women on Boards and Firm Financial Performance: A Meta-Analysis," *Academy of Management Journal* 58, no. 5 (October 2015): 1546–71, https://journals.aom.org/doi/10.5465/amj.2013.0319.

71. Kris Byron and Corinne Post, "Women on Boards of Directors and Corporate Social Performance: A Meta-Analysis," *Corporate Governance: An International Review* 24, no. 4 (July 2016): 428–42, https://onlinelibrary.wiley.com/doi/10.1111/corg.12165.

72. Alice H. Eagly, "When Passionate Advocates Meet Research on Diversity, Does the Honest Broker Stand a Chance?," *Journal of Social Issues* 72, no. 1 (March 2016): 199, https://spssi.onlinelibrary.wiley.com/doi/abs/10.1111/josi.12163.