Colliders in the Boardroom?
The Perils of Collider Bias in Strategy and Management Research

Justin Frake
Andreas Hagemann
Jose Uribe

April 2023

ABSTRACT

In line with published research, we estimate models suggesting that women and minority CEOs can reduce the compensation and representation of other women and minorities, respectively, on a company's top management team (TMT). However, we argue that these correlations are likely due to collider bias, an endogenous selection bias that has not received systematic attention in strategy and management. We use Monte Carlo simulations to illustrate conditions that reduce or amplify the problem and provide generalizable approaches to mitigate the risk of collider bias in applied research. In doing so, we find no evidence that women and minority CEOs damage the career outcomes of other women and minorities in their organizations and highlight the practical threat that collider bias can pose to empirical research.

1 | INTRODUCTION

Strategy and management scholars are increasingly concerned with potential sources of bias in their empirical work. In this regard, significant strides have been made to address the threats to inference posed by omitted variables and non-classical measurement error (Shaver, 1998; Bascle, 2008; Hamilton and Nickerson, 2003; Ge et al., 2016; Wolfolds and Siegel, 2019). However, we contend that collider bias is a significant concern in empirical research that is not yet sufficiently recognized in the management and strategy literature. Collider bias occurs when the relationship between two variables is distorted by conditioning (e.g., controlling or selecting) on a common effect or "collider" variable (Elwert and Winship, 2014; Cinelli, Forney, and Pearl, 2020; Griffith et al., 2020; Schneider, 2020). Despite the potential impact of collider bias on the validity of empirical results, its effect is often overlooked.

We illustrate how collider bias can induce spurious relationships that can have potentially negative practical effects on managerial decisions. An expansive literature examines the effects of having a CEO who is either a woman or a racial minority on firm outcomes (e.g., Cook and Glass, 2014; Jeong and Harrison, 2017; Jeong et al., 2021). In this work, an important human capital outcome for firms is the level of diversity of top management teams and the career advancement of women and minorities (e.g., Dezső et al., 2016; Derks et al., 2016; MacDonald et al., 2018; Chang, et al., 2019; Corwin et al., 2022). We motivate the problem of collider bias with a stylized derivation of the bias term in an examination of the effect of having a woman CEO on the compensation of other women in the top management team (TMT), a subject of recent scholarly attention (cf. Dezső et al., 2022).

Beyond gender, scholars have also shown increased interest in the drivers and consequences of having CEOs who are racial or ethnic minorities (e.g., MacDonald et al., 2018;

Gligor et al., 2021; Jeong et al., 2021). We carry out an empirical demonstration of collider bias using the well-known ExecuComp and Compusat datasets to examine the relationship between women and minority CEOs on both the prevalence and compensation of women and minorities in top management teams from 1992 to 2021. In line with extant work on gender, our analysis shows a large and precisely estimated negative effect of having a woman CEO on the compensation (cf. Dezso et al., 2022) and prevalence (cf. Corwin et al., 2022) of other women and minorities in the top management team. Similar to the case of women CEOs, we show that a CEO who is a racial or ethnic minority can also generate large negative statistical effects on the compensation and representation of minority executives.

Rather than gender or racial dynamics, we argue that the effects we find are a result of conditioning on a collider, namely the probability that women and minorities exit the sample when promoted from non-CEO to CEO positions. To support this claim, we show large and precisely estimated negative relationships between having a CEO with a variety of placebo characteristics (such as having the name "John") on non-CEOs who share the same characteristic.

Using Monte Carlo simulations, we demonstrate how collider bias can induce large spurious correlations. The size of the collider bias in our stylized example is a function of the proportion of women in the TMT, the size of the TMT, and the compensation premium between TMT members who will become CEOs and those who will not. We also discuss the benefits and drawbacks of simple corrections such as using individual fixed effects, removing observations for individuals who will be promoted in the future, inverse probably weighting, multiple imputation, and placebo analysis.

Our study draws attention to the importance of considering collider bias in empirical research in management and strategy. Perhaps the most fundamental lesson is a renewed call for empiricists to attend to the data-generating process. In our example, the data-generating process includes unexpected panel attrition: promotion into the CEO position implies exit from the sample. The well-documented publication pressures on faculty, particularly pre-tenured faculty, and the field's penchant for surprising and counterintuitive findings make collider bias a particularly pressing problem. While other types of empirical problems such as classical measurement error often merely attenuate the effects, collider bias can, and often does, reverse the sign on estimated coefficients. Even though there is no silver bullet for solving collider bias when present, it is critically important to acknowledge its presence. In our example, failure to detect collider bias results in findings that some could use to support policies that reduce the promotion of women and minorities into the CEO position. Broader awareness can therefore increase the validity and usefulness of our findings for informing theory and practice.

2 | BACKGROUND

2.1 | Brief Overview of Collider Bias

Recent trends in the fields of strategy and management have urged scholars to recognize and address potential sources of bias in their empirical findings (Hamilton and Nickerson 2003; Certo et al., 2016; Wolfolds and Siegel, 2019; Stern et al, 2021). One approach to illustrating bias is directed acyclic graphs (DAGs), which were developed by Pearl (2000). This graphical approach uses nodes to represent constructs or variables and arrows to represent causal effects. When a box is drawn around a node, it represents conditioning on the variable (See Figure 1).

We begin by using a DAG to illustrate a confounder in Figure 1A. Most empirical researchers in strategy and management would recognize the need to condition on *M* when investigating the effect of *X* on *Y* to address omitted variable bias.

Less understood, however, is the threat posed by conditioning on colliders. A collider is any variable that is caused by two other variables. Importantly, conditioning on a collider causes rather than ameliorates bias between the two contributing variables. For example, in Figure 1B, *M* is a collider because it is caused by both *X* and *Y*. Whereas conditioning on M in Figure 1A eliminates omitted variable bias when estimating the relationship between *X* and *Y*, conditioning on M in Figure 1B *causes* bias. Importantly, it is possible to cause collider bias between two variables even if they do not directly influence the collider, as shown in Figure 1C. Here, the collider, M, is neither an outcome of X nor Y. However, M is an outcome of U and V, which are antecedents of X and Y respectively. Thus, conditioning on M causes bias when estimating the effect of X on Y.

[Insert Figure 1 here]

Note that by "conditioning" on a collider, we mean controlling, selecting, or stratifying. When a researcher controls for a collider, this is often referred to as a "bad control" (Angrist and Pischke, 2009, p.64; Cinelli, Forney, and Pearl, 2020). Bad controls can generally be avoided by not including controls for intermediate outcomes in the regression.[1] However, collider bias can be more difficult to detect if the collider influences the sample selection process. For this reason, we focus on collider bias from selection.[2]

[1] Controlling for an intermediate outcome is not necessarily a bad control. See Figure 11 of Cinelli et al., (2020) for examples. Moreover, bad controls are not always intermediate outcomes. See Figure 4 of Cinelli et al., (2020).
[2] Note that some confounders can also be colliders (and vice-versa). In fact, in the empirical section, we will return to such a case.

Selection bias is also relatively well understood in strategy and management research. Selection into treatment, whereby unobservable factors drive selection into treatment and affect the outcome variable, is a type of omitted variable bias (see Figure 1A). Sample selection, on the other hand, occurs when nonrandom samples from the population are used to test statistical relationships (Certo, et al., 2016). In this case, the researcher is unable to observe variables for part of the population. Researchers commonly assume that, while sample selection can affect the generalizability of a study's results, it should not strongly influence the observed relationships between variables within the subsample. For example, scholars using samples of publicly traded companies may believe that the sampling approach simply means that their results may not generalize to private companies. While this is true in some cases, this overlooks the possibility that entry into a stock exchange could be a collider, and selection on a collider can cause spurious correlations, even within the selected sample.

To give an example, imagine that a researcher hypothesized that individuals with high technical skills will invest less in their social skills. To test this hypothesis, the researchers took a random sample of the US population and implemented a randomized controlled trial (RCT) that increased the treatment group’s technical skills through a training program that had 100 percent compliance. Then imagine the researchers waited ten years and measured every US worker who was involved in their RCT and precisely measured their social skills. After analyzing the data, the researcher observed a strong and statistically significant negative effect of the training program on subject’s social skills. The researchers then conclude that the results of the RCT support their hypothesis that increased technical skills cause a decrease in social skills. But in our thought experiment, the researchers are wrong. The correlation is an artifact of collider bias. But how can collider bias cause spurious correlations even when the researchers used an RCT with perfect compliance, obtained precise measures of their variables, and sampled the entire U.S. workforce?

The root problem with the above thought experiment arises from the fact that the final sample only comprises employed individuals within the United States and not the entire US population. This implies that the sample is not representative but selected. Although it may be tempting to consider this a minor issue, which merely limits the generalizability of the study's findings to the employed population within the United States (rather than the full U.S. population), this is not the case. Rather, the observed effect of technical skills on social skills is spurious. Even within the sample of US workers, there is no causal effect of technical skills on social skills. This is because the sample is not merely unrepresentative of the population, it is selected on a collider.

To clarify, assume there is no causal connection between technical skills and social skills and that they are independently distributed in the U.S. population (see Panel A in Figure 2). In this population, people with very low technical and low social skills are more likely to be unemployed. Figure 2B demonstrates the negative correlation after removing unemployed people from the population. In this case, the negative correlation results from conditioning on a collider—entry into the U.S. workforce —not any causal connection between the variables. This is represented graphically in the DAG in Figure 3.

To make clear the differences between omitted variable bias, bad controls, and collider bias from selection, we have created a simple app available at https://collider.shinyapps.io/ColliderApp/. This app can also be used by researchers to quantify the direction and magnitude of collider bias in their own studies.

[Insert Figures 2 and 3 here]

2.2 | Collider Bias in Published Papers

The threat of collider bias extends throughout the social sciences (Elwert and Winship, 2014; Schneider, 2020).[3] In economics, a number of published papers have been criticized for the potential presence of collider bias including, among others, Acemoglu, Johnson, and Robinson (2001), Fryer (2019), and the literature on the "early industrial growth puzzle". Acemoglu, Johnson, and Robinson (2001) argue that settler mortality rates at the time of colonization affected whether the settlers established institutions that were inclusive or extractive (e.g., slavery). However, they use post-colonization mortality measures as a proxy for settler mortality at time of colonization. The mortality data come mostly from non-combat deaths of European military in the nineteenth century (Albouy, 2012). Because plantations and the slave trade caused the yellow fever epidemic among European settlers, later mortality measures are heavily influenced by this epidemic. Conditioning on post-colonization mortality is therefore unlikely to recover the relationship between mortality and slavery because post-colonization mortality was endogenously determined by this relationship.

Fryer (2019) studies racial bias in policing using administrative data on police stops. He finds that there are no racial disparities in officer-related shootings. However, if officers have a higher threshold for stopping White[4] versus non-White civilians, then encounters with White civilians in the data pose a higher risk to police than those with non-White civilians (Knox, Lowe, and Mummolo, 2020). This difference in thresholds in the data could therefore mask a much higher likelihood of officers using deadly force against Black and Hispanic people.

The literature on the "early industrial growth puzzle" or "Antebellum puzzle" (Kolmos, 1998) makes the observation that mean heights of adult men fell during the nineteenth century

[3] Elwert and Winship refer to collider bias as "endogenous selection bias"
[4] "White" here refers to non-Hispanic White individuals

while real wages grew. However, Bodenhorn, Guinnane, and Mroz (2017) point out that the available height data come from military and criminal records. Because tall men had better and broader job opportunities over time, shorter-than-average men ended up in the military. The negative correlation between height and real wages is then a result of implicitly conditioning on being in the military.

Relative to economics, much less attention has been paid to collider bias in strategy and management research. We contend that collider bias is at least as relevant to strategy and management scholars as it is to other social sciences. The threat of collider bias potentially exists whenever researchers condition on firm or individual characteristics or choices that could be colliders, such as making part of a market index or having its stock listed on an exchange. For example, Miller, Le Breton-Miller, and Lester (2010) find that family ownership is negatively related to acquisitions amongst publicly traded firms. Yet the results could suffer from collider bias because being listed on an exchange (the potential collider) is likely affected by both the firm's family ownership and by its intention to acquire (because going public raises capital necessary for acquisition).

As another example, several papers have investigated the relationship between patenting and an inventor's propensity to change employers (Hoisl, 2007; Palomeras and Melero, 2010; Melero, Palomeras, and Wehrheim, 2020). Yet, many of these papers condition the sample on inventors who have patented at least twice in their career. This is because they use patent records to measure inventor mobility and need at least two patents to measure one mobility event (one patent to identify the source firm and one to identify the recipient firm). However, if patenting once increases the probability that an inventor will patent again, and if moving firms affects an inventor’s probability of patenting (Kaiser, Kongsted, and Rønde, 2015), then conditioning on the inventor’s total number of career patents is likely to induce collider bias.

2.3 | Stylized illustration

We next illustrate the problem concretely in the context of strategic human capital. We explore the question of the effect of having a woman CEO on the compensation of other women managers in the top management team (TMT) of the firm (cf., Dezso et al., 2022). We will return to this application throughout the remainder of this paper. In this context, the treatment variable is having a woman CEO, and the outcome is change in the compensation of other women in the TMT.

This is a challenging question to answer empirically. A naïve researcher might regress all women TMT members' (including the CEO) compensation on whether there was a woman CEO in the focal or previous year. However, if the sample includes the CEO, then the researcher would wrongly attribute the increased compensation caused by a woman TMT member becoming a CEO to the fact that there is a woman CEO. For example, in Figure 4A, the woman CFO in period 1 was promoted to CEO in period 2 and received an increase in compensation with her promotion. However, if we include the compensation of the woman CEO in the analysis, we would estimate that having a woman CEO increased the compensation of women TMT members by $25,000, even though there was no causal effect of having a woman CEO.

[Insert Figure 4 here]

To address this problem, the authors exclude CEOs from their sample and condition the analysis on non-CEO women only, as Seen in Figure 4B. While this approach addresses the previous problem, it also introduces collider bias. Here, we would wrongly estimate that women CEOs cause TMT women to experience a \$25,000 *reduction* in their wages.

We represent the problem of collider bias using a DAG in Figure 5. *Internal promotion* of a woman in the TMT causes the TMT member's rank to change (e.g., from CFO to CEO) and causes there to be a woman CEO. Further, a TMT member's *human capital* (which may include experience, social connections and other factors) affects their *rank* (TMT members with greater human capital are more likely to be CEO) and their *compensation*. The DAG makes it clear that the TMT member's rank is a collider on the path from *woman CEO* to *compensation*. Thus, by excluding CEOs from the analysis, the researchers are conditioning on rank, which is a collider.

[Insert Figure 5 here]

Next, we generalize and formalize the intuition in this example by deriving the bias term using a simple model. As in Figure 4, suppose there are two time periods $t = 1,2$ and two women $i = 1,2$ with compensation $Y_{i,t}$. In period 1, both women are on the TMT in non-CEO positions but in period 2 the individual with greater human capital of the two women gets promoted to CEO. Suppose higher human capital manifests itself in higher compensation in period 1 so that $Y_{2,1} = Y_{1,1} + W$, where the compensation gap W is a random variable with $\delta = E(W) > 0$.

The researcher uses a stylized regression model with firm fixed effects to account for unobserved firm characteristics (e.g., a women-friendly environment). We capture this in the model for average women's TMT salary:

$$\bar{Y}_t = \beta_0 + \beta_1 1\{\text{woman CEO at time } t\} + U_t + V. \qquad (1)$$

The term U_t is a mean-zero idiosyncratic error uncorrelated with all other variables and V is a firm-specific unobserved variable potentially correlated with whether the firm has a woman CEO. To remove V from the equation, the researcher uses fixed effects estimation (which is equivalent to first differences here) to estimate:

$$\Delta\bar{Y} = \beta_1 + \Delta U_t, \tag{2}$$

where $\Delta\bar{Y} = \bar{Y}_2 - \bar{Y}_1$, β_0 and V have dropped out because they are time-invariant, and β_1 remains because a woman transitioned into the CEO role. Taking expectations yields that

$$\begin{aligned} \beta_1 &= E(\Delta\bar{Y}) = E\left(Y_{1,2} - \frac{Y_{1,1}+Y_{2,1}}{2}\right) = E\left(Y_{1,2} - \frac{2Y_{1,1}+W}{2}\right) \\ &= \underbrace{E\left(Y_{1,2} - Y_{1,1}\right)}_{\text{average change in compensation}} - \underbrace{\delta/2}_{\text{collider bias}} . \end{aligned} \tag{3}$$

In words, the regression coefficient β_1 measures the average change in the compensation of the remaining woman employee minus a bias term that measures the compensation differential between the promoted woman and the remaining woman. If this differential is large, then the bias term will overwhelm even a large *increase* in women's compensation. A regression of average non-CEO women's compensation on indicators for having a woman CEO will have a non-ignorable and systematic downward bias. This bias will not be present in regressions of the effect of having a woman CEO on men's compensation. Large negative coefficients are not informative, and at best, lower bounds.

The 2 in the bias term ($\delta/2$) stands for the number of employees, so bias can be expected to be smaller if the number of non-CEO women on the TMT is large (unless having more women makes the compensation differential W larger). Formalizing this statement makes the derivation more complex but the same idea applies. Suppose there are n women at time $t = 1$.

Compensation packages $Y_{1,t}, \dots, Y_{n-1,t}$ are identically distributed copies of a random variable Y_t. The compensation $Y_{n,t}$ of the n-th person (the future CEO) is larger on average, such that $E(Y_{n,t}) = E(Y_t) + \delta$. Let $\bar{Y}_1$ be the average with all observations at time $t = 1$, $\bar{Y}_1^*$ be the average with observation n removed at time $t = 1$, and $\bar{Y}_2^*$ be the average with observation n removed at time $t = 2$. If observation n becomes CEO, the fixed effects regression coefficient now identifies

$$\begin{aligned}\beta_1 &= E(\Delta\bar{Y}) = E(\bar{Y}_2^* - \bar{Y}_1) = E\left(\bar{Y}_2^* - \bar{Y}_1^* + \frac{\bar{Y}_1^* - Y_{n,1}}{n}\right)\\ &= \underbrace{E(\bar{Y}_2^* - \bar{Y}_1^*)}_{\text{average change in compensation for non-CEOs}} \quad \underbrace{- \delta/n}_{\text{bias}}.\end{aligned} \tag{4}$$

The δ/n term is a manifestation of collider bias. We illustrate the mechanics graphically in Figure 6, which plots compensation against whether there is a woman CEO. The black dots represent observations of non-CEO women on the TMT, and the red dots indicate the counterfactual compensation of the women CEOs had they not been promoted, but still had a woman CEO. Thus, we have constructed the data such that there is no causal effect of CEO gender on the compensation of non-CEO women. This is represented by the flat line in Figure 3A. However, Figure 3B demonstrates that the correlation is negative once the women who were promoted to CEO are removed. The fact that we systematically do not observe women who were promoted to CEO in period two acts as the collider.

[Insert Figure 6 here]

3 | EMPIRICAL ILLUSTRATION

Until now, we have used thought experiments, mathematical models, and simulated data to illustrate the problem as simply as possible. In this section, we demonstrate how collider bias can result in spurious findings using real-world data to address a question that is relevant to strategy and management scholars. Specifically, we produce results that are consistent with findings in Dezso et al. (2022), who find that the presence of a woman CEO (compared to a man CEO) reduces the compensation of other women on the TMT by over 16%. Based on these findings, the authors argue that having a woman CEO reduces the diversity benefits contributed by other women in the TMT. Relatedly, Corwin et al (2022) find that, compared to a man CEO, the presence of a woman CEO reduces the number of other women in the TMT by 7% in the subsequent year. Based on these results, the authors argue that women CEOs may actively exert pressure to resist the advancement of other women in the company (the so-called "queen bee" effect). These are not inconsequential findings, especially if policymakers use them as the basis for decision-making. For example, one possible implication of these findings is that if boards want to reduce the gender pay gap or the "glass ceiling," they may hesitate appointing women to the CEO position.

We investigate these findings in the remainder of this section from the perspective of collider bias. To summarize, we use standard OLS regressions with firm and year fixed effects to estimate the effects of women CEOs on other women managers. Our results are similar in magnitude and precision to those reported by Dezso et al. (2022) and Corwin et al (2022). To demonstrate how scholars could extend these findings using other CEO characteristics, we replicate the same analysis for racial and ethnic minorities. Here again, the results suggest that the presence of a minority (e.g., non-white) CEO dramatically reduces the compensation of other minorities on the top management team, and has a milder but still negative effect on the proportion of other minorities on the TMT in subsequent years.

While in line with the findings in extant work, we aim to demonstrate that the observed correlations we report are unrelated to the gender or minority status of the CEO and can be explained by collider bias. Again, transitioning from a non-CEO to a CEO position is the collider, and by conditioning the sample on individuals who are not CEOs, these analyses are threatened by collider bias. To substantiate our claim, we use placebo regressions (e.g., Jarosiewicz & Ross, 2022) to show that various CEO characteristics unrelated to gender and minority status produce comparable effects on the outcomes of non-CEOs sharing the same characteristic. In a later section, we show that excluding observations from individuals who transition between non-CEO and CEO positions eliminates the effects across the board.

3. 1 | Data and sample[5]

We use 30 years of data (1992 – 2021) from the Compustat and ExecuComp datasets to examine the effect of having a woman or minority CEO on two important outcomes covered by prior research: 1) the compensation of other managers in the TMT (cf. Dezso et al., 2022) and 2) the proportion of other women in the TMT (cf., Corwin et al., 2022).[6] We select firms in the S&P 1500 index and all employees in ExecuComp with valid compensation data. As in Dezso et al. (2022), we construct the TMT as consisting of the five highest-paid employees, namely the CEO and four additional employees holding titles such as COO, CFO, or Executive VP. Our final sample consists of 143,236 observations for 25,161 executives working at 1,495 firms. Observations for women employees constitute 7.90% of the total sample, while racial/ethnic minorities constitute 7.15% of the sample.

[5] We invite readers to reproduce our results using the annotated Stata code available with this article.

[6] While we follow the same general empirical approach as these authors, there are some differences in our models. First, our control variables are similar but not identical those used by either set of authors. Second, we use a longer time period: Corwin et al., analyze data from 2011-2017, while Dezso et al, analyze data from 1992 – 2013. We use 1992 – 2021. Third, Corwin et al.(2022) conduct a first-stage probit regression predicting the presence of a woman CEO in the firm. This step does not address collider bias, but can address firms' self-selection into the treatment condition, as long as the first-stage regression contains an exogenous instrument (see Wolfolds & Siegel, 2019).

The ExecuComp dataset includes a variable indicating whether an executive is a man or a woman. The race and ethnicity of TMT members are, however, not readily available. But McDonald et al., (2018) report that 8% of executives in large and medium U.S. based companies are racial minorities. ICS, a consultancy, breaks down this percentage for the Russell 3000 into Asian (5%), Hispanic (2%) and Black (1%).[7] Using this breakdown, we categorized executives as racial/ethnic minorities based on their first and last names and zip code with NamSor's machine learning algorithm. NamSor uses a probability distribution of names to estimate their race and ethnic origin (for details, see Santamaría & Mihaljević, 2018). An executive was therefore categorized as a minority if their name was ranked at the top 5%, 2% or 1% of Asian, Hispanic or Black names, respectively. This categorization of minorities is obviously noisy and imprecise, but adequate to demonstrate the effects of collider bias.

3.2 | Placebo Groups

Placebo groups are unrelated to gender or minority status. Regression results from placebo groups, therefore, reflect the general statistical properties of the data-generating process. Thus, if placebo regressions exhibit similar correlations to those found for women and minorities, we can reasonably infer that the correlations are not driven by the authors' proposed mechanisms, but rather by some other factor.

We compare results for women and minorities with identical regressions for three placebo groups of employees that appear in the dataset at comparable rates. We constructed the following three groups: employees whose first name is "John" (4.83% of the total sample), employees whose first name starts with the letter "M" (9.80% of the sample), and a group of

[7] see https://www.isscorporatesolutions.com/library/data-snapshot-board-and-workforce-racial-diversity/ (last accessed 2/28/2023)

randomly selected employees (7.71% of the sample). The number of observations in the random group is slightly greater than the proportion of minorities and slightly less than the proportion of women.[8]

Women and minorities in our sample are largely non-overlapping with each other or with the placebo groups: 10% of women are also categorized as minorities, while 11% of minorities are also categorized as women. The "John" group contains 0% women and 2% minority observations, the "Letter M" group contains 10% women and 6% minority observations, and the "Random" group contains 8% women and 7% minority observations. Table 1 below describes the size of the various groups used in the analyses.

[Insert Table 1 here]

3.3 | Variables

Top Manager compensation (Dependent Variable 1). Following Dezso et al., (2022) we used the natural log transformation of a top manager's total compensation, including salary, bonus, and grants of stock and options.

TMT Representation (Dependent Variable 2). Following Corwin et al., (2022), we calculate *% Women in TMT* as the percentage of women TMT members (excluding the CEO) at time t + 1. A measure of 0 indicates that there are no women on the TMT, while a measure of 1 indicates that every member of the TMT is a woman. We construct parallel measures for *% Minority in TMT, % "John" in TMT, % "Letter M" in TMT* and *% Random Group in TMT.*

Has CEO of Type X. Following recent work (Corwin et., 2022; Dezso et al., 2022) *Woman CEO* takes the value 1 if, in a given year, a firm has a woman CEO and 0 otherwise. We construct parallel measures for *Minority CEO, John CEO, Letter M CEO* and *Random Group CEO.*

[8] Results are robust to a very wide range of randomly chosen groups.

Controls: We used a similar suite of firm-level and executive-level control variables with those in extant work (see Table 2 in Dezso et al., 2022 for explanations). At the firm level, these include logged transformations of *Book Leverage, Advertising intensity, R&D intensity, size from assets, size from employees, Age of Capital Stock* and *Tobin's Q.* At the individual executive level, we included a log transformation of the *Executive's age,* and whether the Employee holds the title of CFO (*Employee is CFO*).

Table 2 below contains descriptive statistics and zero order correlations for the variables in our analyses.

[Insert Table 2 here]

3.4 | Results

3.4.1 | The Effect of Woman CEOs on the Compensation of Other Top Managers

We use two-way fixed effects for firms and years to estimate the coefficient of having a woman or minority CEO on the total compensation of other women or minorities on the TMT. As in Dezso et al. (2022), we run the analysis in the unpooled sample corresponding to the group sharing the same characteristic as the CEO.[9]

Model 1 in Table 3 shows a 17% decrease in the compensation of other women in the TMT from having a woman CEO compared to a man CEO. This effect is very similar in size to the effect reported by past research (cf. Table 5, Model 2 in Dezso et al., 2022). In Table 3, Model 2 we also find that, compared to having a white CEO, having a minority CEO reduces the compensation of other minorities by nearly 29%. These regressions do not, however,

[9] Deszo et al (2022) support their argument by showing a null effect for having a woman CEO on the compensation of men. These results are available upon request, but we do not include them here because they are irrelevant to our demonstration of collider bias. This is because in the ExecuComp dataset non-CEO men/non-minorities do not exit the sample to become women/minority CEOs.

demonstrate that collider bias is driving the effects. It is possible that women and minority CEOs do have real negative effects on other women and minority TMT members. To explore whether this is the case, we turn to placebo regressions.

The results are quite similar for the placebo groups. Individuals named “John” (Model 3) appear to suffer a 24% decrease in their compensation when their CEO is also named John, compared to having a CEO with another name. We also find large, and precisely estimated negative effects on the compensation of executives whose name starts with the letter “M” or who belong to a random group, from having a CEO sharing that specific trait. The results cast doubt on the veracity of claims that compared to men CEOs, women CEOs hurt the compensation of other women TMT members.

3.4.2 | The Effect of Woman and Minority CEOs on TMT Gender and Ethnic Diversity

Next, we follow Corwin et al (2022) and collapse the data into firm-year observations to analyze the effect of having a woman or minority CEO on the proportion of other women and minorities on the TMT in the subsequent year. Table 4 displays the coefficients for the five dependent variables.

[Insert Tables 3 and 4 here]

The magnitude of the effect of a woman CEO on the proportion of women on the TMT in the subsequent period is -3.7%, about half of the negative effect reported by Corwin et al., (2022). The effect of women CEOs on gender representation is very similar to the effect of CEOs from the random group on the proportion of employees from the same random group (-3%, see model 5 in Table 4). The CEO effect is smaller for minorities and employees named “John” (-1.6%), and larger for employees whose name starts with the letter “M” (-4.1%). Across all groups, the overall pattern produced by collider bias is consistent. Unless we are to believe that there are causal mechanisms causing CEOs named John to push out other TMT members named John, the results again cast doubt on the causal explanation that women and minority CEOs erode the gender and monitory diversity of top management teams.

3.4 | Correcting Collider Bias

3.4.1 | Excluding Future CEOs

The solutions to collider bias vary by context. The simplest solution is to avoid conditioning on colliders. Yet, sometimes this solution is not available. For instance, in our example in section 2.3, both including and excluding the observations of women CEOs can cause bias. In this case, the best solution may be to exclude observations that may cause collider bias. For our empirical example, this is done by excluding all women who were *ever* CEO, not just the years that they were CEO. This approach is very similar to including individual-level fixed effects because there is no within-person variation in *WomanCEO* for women who are internally promoted to CEO. In Tables 5 and 6 we compare the coefficients and standard errors previously reported in Tables 3 and 4 with those of identical regressions run on samples that exclude individuals who will become CEOs in future periods.

Table 5 illustrates the dramatic changes in non-CEO compensation resulting from removing a small percentage of observations that transition from the non-CEO sample to the (excluded) CEO sample. The large and precisely estimated negative coefficients for women, minorities, and the three placebo groups become smaller and statistically indistinguishable from zero.

[Insert Table 5 here]

Similarly, Table 6 illustrates the changes in TMT representation resulting from removing a small percentage of observations that transition from the TMT sample to the (excluded) CEO sample.

The coefficients become smaller and less negative or even positive for some groups. In particular, after correcting for collider bias, the coefficient suggests that minority CEOs may *increase* the representation of other minorities in the TMT in the subsequent year.

[Insert Table 6 here]

Including individual fixed effects or removing "transitioning" observations both eliminate collider bias in this application. However, this does not mean that the resulting estimates are unbiased for the entire population of TMT members. By excluding women who ever become CEO from the sample, the estimates are only valid for executives who never transition from the non-CEO to the CEO pool, rather than all executives.

This empirical example demonstrates how conditioning on colliders can severely bias empirical analysis, resulting in potentially spurious findings. We can reasonably assume that the true effect in the three placebo groups ("John", "First Letter M" and "Random") is zero. A natural question then is what can explain the wide range in our spurious findings, from -24% to -11% in Table 5 and from -4% to -1.6% in Table 6. Next, we use Monte Carlo analysis to explore the conditions under which the bias is more or less severe.

4 | MONTE CARLO ANALYSIS

Monte Carlo simulations allow researchers to define a data-generating process, which is a set of assumptions about how data are generated from underlying population parameters. By defining the data-generating process, researchers can "know" the "true" causal effect because they are responsible for defining the data generating process. After defining the relationship between variables, the researcher then simulates data that reflects the assumed relationships between variables. By running regressions on the simulated data, researchers can then test whether their empirical approach and/or modeling choices are able to recover the correct coefficients. This allows researchers to evaluate the performance of statistical methods and models under different scenarios, and to identify potential sources of bias that may affect the accuracy of their results. In the following, we use Monte Carlo simulations to investigate the sensitivity of the collider bias observed in the preceding section to the following key attributes: the proportion of women in the TMT, the size of the TMT, and the compensation premium that future CEOs experience prior to being promoted to CEO.

4.1 | Baseline Setup

For each scenario, we simulate 100 datasets with 1,000 observations each according to the following data-generating process:

$$W_{i,k,t} = (H_{i,k} + \varphi_{i,k,t})\ \lambda_i\ C_i \tag{5}$$

where $W_{i,k,t}$ represents individual i's compensation in firm k in period t. The term $H_{i,k}$ represents the TMT member's human capital at time t. The term $\varphi_{i,k,t}$ is an random error term distributed normally with a mean of five and a standard deviation of one and represents all of the other factors that influence compensation other than human capital. The term λ_i represents the compensation premium, in percentage terms, that individuals who eventually become CEOs receive relative to the average compensation of other TMT members. C_i is a dummy equal to one if individual i ever becomes CEO. Note that, because there is no term for the gender of the individual or CEO in equation 5, this data-generating process assumes that compensation is independent of these factors. Thus, the "true" causal effect of a woman CEO on others' compensation as defined by the data generating process is zero. Our baseline simulation assumes just two periods (t=1,2). We also assume that the compensation gap (λ) is 10 percent, each TMT consists of five people, and that 20 percent (on average) of TMT members are women. We relax these assumptions sequentially to investigate their effect on the bias.

4.2 | Monte Carlo Results

In Figure 7, we vary λ_i, the compensation premium that future CEOs receive, from 0 to 50 percent above the average of other TMT members. The results demonstrate that, in line with equation 4 in section 2.3, the estimates are biased when TMT members who are eventually promoted to CEO are paid even 5% more than the average of other TMT. This bias increases approximately linearly with the compensation premium of future CEOs.

[Insert Figure 7 here]

Having shown how differences in the compensation gap (the numerator of equation 4 in section 2.3) affect the results, we turn to factors that affect the denominator of equation 4 in section 2.3, which is the number of non-CEO women on the TMT. Specifically, the number of non-CEO women on the TMT is likely to increase as the size of the TMT increases and as the proportion of women in the TMT increases.

In Figure 8, we set the compensation gap equal to 10 percent and vary the top management team size from five to 100 in increments of five. The results demonstrate that as the size of the top management team increases, the bias reduces logarithmically.

[Insert Figure 8 here]

In Figure 9, we again fix the compensation gap at 10 percent and the number of non-CEO TMT members at five. We vary the proportion of women in top management teams from 20 percent to 80 percent. The results show that as the proportion of women on the top management team increases, the bias decreases somewhat but remains strongly negative. The intuition behind this result is that as there are more women on the top management team, a single woman being promoted to CEO has a smaller impact on the mean compensation of other women in the second period.

[Insert Figure 9 here]

4.3 | Solutions to Collider Bias

There are several strategies to address collider bias in observational studies. No single solution is universally applicable. Rather, the right approach depends on the study design and data available. In what follows, we present a series of practical approaches and solutions to collider bias.

DAGs. The first step to eliminating collider bias is recognizing its existence. Thus, we encourage researchers to use Directed Acyclic Graphs (DAGs) to understand whether collider bias is a problem in their setting. While a complete treatment of DAGs is beyond the scope of this paper, we point the interested reader to Chapter 6 of Huntington-Klein (2022). Briefly, researchers should first identify the treatment, outcome, and other relevant variables that cause or are caused by the treatment and outcome. Connect the variables (nodes) with arrows (directed edges) to represent the causal relationships between the variables. Ensure that the arrows point from the cause to the effect and that there are no feedback loops. Once the DAG is complete, researchers can examine the DAG to identify colliders, which are variables that are the common effects of two or more other variables. A collider can be recognized by having two or more arrows pointing to it.

Avoiding "bad controls." After understanding the causal relationships and potential colliders using the DAG, researchers should avoid conditioning on potential colliders. In regression analysis, this means excluding controls that might be colliders (Cinelli et al., 2020). As a general heuristic, we recommend avoiding controlling for any potential intermediate outcomes, or variables that were measured after treatment.

Fixed effects and subsampling. While the solution to "bad controls" is to not control for them, the solutions to selection on colliders are less straightforward. This is because, in many cases, a sample that is not selected on a collider is not available. In this case, there are several potential solutions. One potential approach is to subsample on observations that do not suffer from collider bias. In our case, this would involve removing all observations for individuals who ever became CEO or including individual fixed effects, which are roughly equivalent in our setting. For example, either of these approaches would eliminate the bias illustrated in Figure 7, helping the researcher recover an unbiased estimate of the true causal effect. While this is a reasonable solution in our case, it may not generalize to all cases of collider bias. This is because collider bias often does not only affect a clearly defined set of units (e.g., women who ever become CEO), but may affect all or most units in the sample. This approach can also cause selection bias if the resulting sample is no longer representative of the population.

Inverse probability weighting. An alternative approach to addressing collider bias is through inverse probability weighting (IPW). This method involves weighting observations according to their likelihood of being included in the sample. The objective of this weighting is to balance the representation of units that may be overrepresented or underrepresented as a result of conditioning on the collider. In practice, these weights signify the probability of different units being selected into the sample based on their observable characteristics. For instance, in an empirical example involving CEOs, their selection into the sample might depend on their human capital (as illustrated by the DAG in Figure 5). If a researcher has a proxy measure for human capital, they could use IPW to estimate and adjust for the likelihood of a top management team

(TMT) member being promoted to CEO. Accordingly, we weighted each observation according to the individual's human capital ($1/H_{i,k}$) to correct for their probability of being promoted. As demonstrated in Figure 6, this approach also recovers the true causal effect. Breen and Ermisch (2021) demonstrate that inverse-probability weighting can recover unbiased estimates in the case where selection is a function of the outcome variable only. In other cases, IPW can reduce the bias, sometimes to negligible levels, if certain conditions are met (Griffith et al., 2020).

Multiple imputation. Multiple imputation is another statistical technique that can be employed to address collider bias when the collider is related to missing data. The method involves generating multiple complete datasets by imputing missing values using a suitable model that accounts for the relationships between variables. Each of these completed datasets is then analyzed independently, and the results are combined to produce a single, pooled estimate. By accounting for the uncertainty associated with the imputed values, multiple imputation mitigates the bias introduced by conditioning on the collider while preserving the relationships between the exposure, outcome, and any confounders. In the case of our empirical example, we imputed the wages of women in the years that they were CEOs based on their human capital. As seen in Figure 10, this approach recovers unbiased estimates of the causal effect. It is crucial to note that the effectiveness of multiple imputation in addressing collider bias hinges on the proper specification of the imputation model, which again can be clarified by using a DAG.

[Insert Figure 10 here]

Placebo analysis. In the case where the above approaches are not feasible, researchers may use placebo analysis to explore the likelihood of collider bias in their setting. While this approach cannot necessarily rule out collider bias, it can be useful in identifying cases where collider bias is likely present. To implement a placebo test, the researcher should choose a variable that is unrelated to both the treatment and outcome. Then they include the placebo in the regression alongside the original treatment variable. If the regression estimates a significant effect of the placebo on the outcome when conditioned on the collider, collider bias may be present. This is because any observed relationship between the placebo variable and the outcome is likely due to the bias introduced by conditioning on the collider. To further evaluate the presence of collider bias, the researcher could compare the results of the analysis with and without conditioning on the collider (if possible). A significant difference in the estimates for the exposure or the placebo variable supports the presence of collider bias.

4 | DISCUSSION AND CONCLUSION

Collider bias is a pervasive problem in social science research. Unlike confounders, however, very little scholarship in management and strategy discusses the threats that collider bias presents to the validity of empirical findings. We describe the problem in general terms, provide relevant examples where collider bias may be present in published results, and demonstrate that theoretically and practically important research questions examined in recent research may suffer from collider bias.

Our study on the relationship between the outcomes of women and minority non-CEOs from having a CEO who is a woman or minority provides a clear example of how collider bias can result in spurious findings. Specifically, we obtain large and precisely estimated negative coefficients on the effect of having a woman CEO or a minority CEO on the compensation and representation of other women and minorities in the top management teams of S&P 1500 firms. Our results for women CEOs are similar in terms of magnitude and statistical significance to recent work (i.e., Corwin et al., 2022; Dezso et al., 2022). We provide evidence that collider bias – rather than the gender or minority status of individuals – drives our results. Without awareness of the collider bias problem, we would wrongly conclude that a woman CEO or minority CEO reduces the compensation and prevalence of other women and minorities in the TMT.

We offer several potential remediating approaches. First, scholars can use common statistical adjustments, like fixed effects at the appropriate level, inverse probability weighting, or multiple imputation to reduce collider bias. These solutions may not be a silver bullet. For instance, having a future CEO in the TMT may have influenced the composition of the TMT or the trajectory of the company in ways that cannot be controlled for by removing CEOs from the sample. Removing CEO observations (both current and future CEOs) also changes the composition of the sample such that, at best, the effects recovered from having a woman CEO apply only to TMT members who never become CEOs. While removing collider bias may be difficult, detecting its presence is much simpler. At a minimum, scholars can use DAGs, placebo regressions, and simple Monte Carlo simulations to understand the potential threat of collider bias. For example, researchers studying the presence of women in TMTs may compare the observed distribution across firms with simulated distributions (e.g., Dezso et al., 2016). An examination of the effect of female CEO on broader within-firm gender equality may require obtaining data on gender disparities among employees below the TMT. This would reduce the effect of collider bias caused by attrition due to promotion to CEO.

Strategy and management research has made great strides in increasing empirical rigor within the field (Shaver, 1998; Hamilton and Nickerson, 2003; Ethiraj, Ethiraj, Gambardella and Helfat, 2016; Gambardella and Helfat, 2017; Wolfolds and Siegel, 2019; Quigley et al., 2023). We hope to contribute to this agenda by highlighting the threat of collider bias to the validity of empirical results. While some other empirical problems, like classical measurement error on the independent variable, merely attenuate effects, collider bias can, and often does, result in correlations with the wrong sign. This makes awareness of collider bias particularly important. We hope that by bringing a broader awareness of collider bias, and by providing approaches to mitigate its effects, we can increase the validity and usefulness of our findings for informing theory and practice.

REFERENCES

Acemoglu, D., Johnson, S., and Robinson, J. (2001). The colonial origins of comparative development: An empirical investigation. The American Economic Review, 91(5):1369– 1401.

Angrist, J. D., & Pischke, J. S. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton university press.

Albouy, D. Y. (2012). The Colonial Origins of Comparative Development: An Empirical Investigation: Comment. The American Economic Review, 102(6):3059–3076.

Bascle, G. (2008). Controlling for endogeneity with instrumental variables in strategic management research. Strategic organization, 6(3), 285-327.

Bodenhorn, H., Guinnane, T. W., and Mroz, T. A. (2017). Sample-Selection Biases and the Industrialization Puzzle. Journal of Economic History, 77(1):171–207.

Breen, R., & Ermisch, J. (2021). Using Inverse Probability Weighting to Address Post-Outcome Collider Bias. Sociological Methods & Research, 00491241211043131.

Certo, S. T., Busenbark, J. R., Woo, H. S., & Semadeni, M. (2016). Sample selection bias and Heckman models in strategic management research. Strategic Management Journal, 37(13), 2639-2657.

Chang, E. H., K. L. Milkman, D. Chugh, and M. Akinola 2019 "Diversity thresholds: How social norms, visibility, and scrutiny relate to group composition." Academy of Management Journal, 62: 144-171.

Cook, A. and Glass, C. (2014). Above the glass ceiling: When are women and racial/ethnic minorities promoted to CEO? Strategic Management Journal, 35: 1080-1089.

Cinelli, C., Forney, A., & Pearl, J. (2020). A crash course in good and bad controls. Sociological Methods & Research, 00491241221099552.

Corwin, E. S., Loncarich, H., & Ridge, J. W. (2022). What's it like inside the hive? Managerial discretion drives TMT gender diversity of women-led firms. Journal of Management, 48(4), 1003-1034.

Derks, B., C. Van Laar, and N. Ellemers
2016 "The queen bee phenomenon: Why women leaders distance themselves from junior women." Leadership Quarterly, 27: 456-469.

Dezső, C. L., D. G. Ross, and J. Uribe. (2016) “Is there an implicit quota on women in top management? A large-sample statistical analysis.” Strategic Management Journal, 371: 98–115.

Dezső, C. L., Li, Y., & Ross, D. G. (2022). Female CEOs and the compensation of other top managers. Journal of Applied Psychology, 107(12): 2306–2318.

Elwert, F. and Winship, C. (2014). Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable. Annual Review of Sociology, 40(1):31–53.

Ethiraj, S. K., Gambardella, A., & Helfat, C. E. (2016). Replication in strategic management. Strategic Management Journal, 37(11), 2191–2192.

Ethiraj, S. K., Gambardella, A., & Helfat, C. E. (2017). Improving data availability: A new SMJ initiative. Strategic Management Journal, 38(11), 2145–2146.

Fryer, Roland G. 2019. “An Empirical Analysis of Racial Differences in Police Use of Force.” Journal of Political Economy, 127 (3): 1210–61.

Griffith, G.J., Morris, T.T., Tudball, M.J., Herbert, A., Mancano, G., Pike, L., Sharp, G.C., Sterne, J., Palmer, T.M., Davey Smith, G. and Tilling, K. (2020). Collider bias undermines our understanding of COVID-19 disease risk and severity. Nature communications, 11(1), 5749.

Ge, C., Huang, K. W., & Png, I. P. (2016). Engineer/scientist careers: Patents, online profiles, and misclassification bias. Strategic Management Journal, 37(1), 232-253.

Gligor, D. M., Novicevic, M., Feizabadi, J., & Stapleton, A. (2021). Examining investor reactions to appointments of Black top management executives and CEOs. Strategic Management Journal, 42(10), 1939-1959.

Hamilton, B. H., & Nickerson, J. A. (2003). Correcting for endogeneity in strategic management research. Strategic organization, 1(1), 51-78.

Hoisl, K. (2007). Tracing mobile inventors—the causality between inventor mobility and inventor productivity. Research Policy, 36(5), 619-636.

Huntington-Klein, N. (2022). The effect: An introduction to research design and causality. Chapman and Hall/CRC.

Jarosiewicz, V. E., & Ross, D. G. (2020). Revisiting managerial “style”: The replicability and falsifiability of manager fixed effects for firm policies. Strategic Management Journal.

Jeong, S. H., & Harrison, D. A. (2017). Glass breaking, strategy making, and value creating: Meta-analytic outcomes of women as CEOs and TMT members. Academy of Management Journal, 60(4), 1219-1252.

Jeong, S. H., Mooney, A., Zhang, Y., & Quigley, T. J. (2021) How do investors really react to the appointment of Black CEOs? A comment on Gligor et al. Strategic Management Journal, 1-20.

Kaiser, U., Kongsted, H. C., & Rønde, T. (2015). Does the mobility of R&D labor increase innovation? Journal of Economic Behavior & Organization, 110, 91-105.

Knox, D., Lowe, W., & Mummolo, J. (2020). Administrative Records Mask Racially Biased Policing. American Political Science Review, 114(3), 619-637.

Komlos, John, Shrinking in a Growing Economy? The Mystery of Physical Stature during the Industrial Revolution. The Journal of Economic History, Vol. 58, No. 3 (Sep., 1998), pp. 779-802

Kuhn, T. S. (2012). The structure of scientific revolutions. University of Chicago press.

McDonald, M. L., Keeves, G. D., & Westphal, J. D. (2018). One step forward, one step back: White male top manager organizational identification and helping behavior toward other executives following the appointment of a female or racial minority CEO. Academy of Management Journal, 61(2), 405-439.

Miller, D., Le Breton-Miller, I., & Lester, R. H. (2010). Family ownership and acquisition behavior in publicly-traded companies. Strategic management journal, 31(2), 201-223.

Melero, E., Palomeras, N., & Wehrheim, D. (2020). The effect of patent protection on inventor mobility. Management Science, 66(12), 5485-5504.

Palomeras, N., & Melero, E. (2010). Markets for inventors: Learning-by-hiring as a driver of mobility. Management Science, 56(5), 881-895.

Pearl, J. (2000). Causality. Cambridge university press.

Quigley, T. J., Hill, A. D., Blake, A., & Petrenko, O. (2023). Improving Our Field Through Code and Data Sharing. Journal of Management, 49(3), 875–880.

Santamaría, L., & Mihaljević, H. 2018. Comparison and benchmark of name-to-gender inference services. Peer J Computer Science, 4: e156.

Santamaría, L., & Mihaljević, H. 2018. Comparison and benchmark of name-to-gender inference services. PeerJ Computer Science, 4: e156.

Schneider, Eric B., 'Collider Bias in Economic History Research', Explorations in Economic History, 78, no. 1 (2020)

Shaver, J. M. (1998). Accounting for endogeneity when assessing strategy performance: Does entry mode choice affect FDI survival?. Management science, 44(4), 571-585.

Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. Royal Society open science, 3(9), 160384.

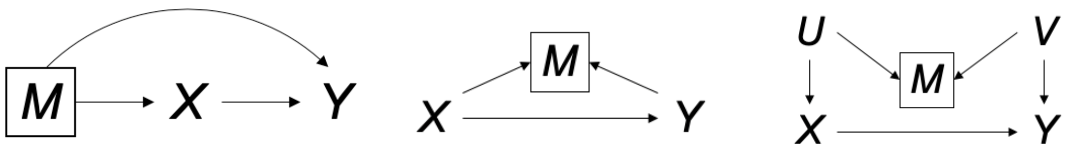
Smith, N., and P. Parrotta. (2018). "Why so Few Women on Boards of Directors? Empirical Evidence from Danish Companies in 1998-2010." Journal of Business Ethics, 147: 445-467.

Stern, I., Deng, X., Chen, G., & Gao, H. (2021). The "butterfly effect" in strategic human capital: Mitigating the endogeneity concern about the relationship between turnover and performance. Strategic Management Journal, 42(13), 2493-2510.

Waddock, S. A., & Graves, S. B. (1997). The corporate social performance–financial performance link. Strategic management journal, 18(4), 303-319.

Wolfolds, S. E., & Siegel, J. (2019). Misaccounting for endogeneity: The peril of relying on the Heckman two-step method without a valid instrument. Strategic Management Journal, 40(3), 432-462.

(A) Confounder **(B) Simple Collider** **(C) M-bias Collider**



M is a confounder. Failing to condition on M causes omitted variable bias.

M is a collider. Conditioning on M causes collider bias.

M is a collider. Even though M is not an outcome of X, it is a collider on the path from X to Y. Conditioning on M causes collider bias.

Figure 1. Illustrating Confounders and Colliders using DAGs

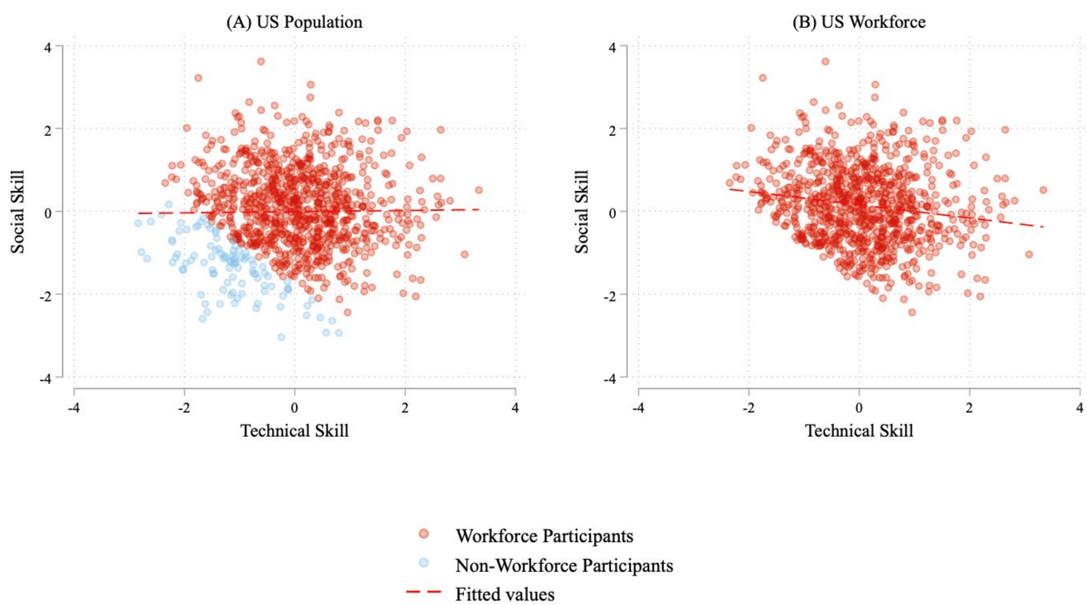


Figure 2: Illustration of Collider Bias in Technical and Social Skill

Notes. Panel A is a scatter plot of 1000 simulated observations where social and technical skills are uncorrelated. Panel B is the same dataset with the most and least skillful observations removed.

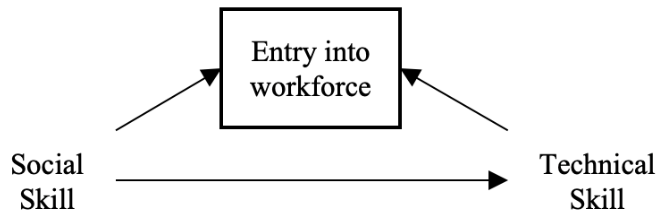


Figure 3: DAG of Collider Bias.

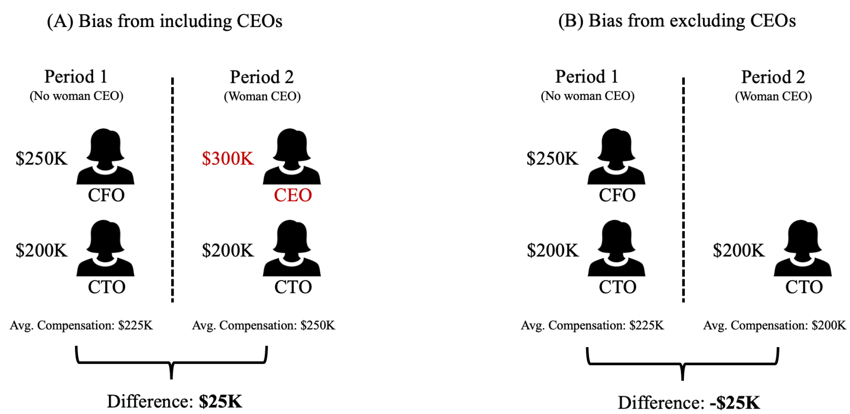


Figure 4: Illustrative example of the challenge of estimating the effect of a Woman CEO on TMT women's compensation.

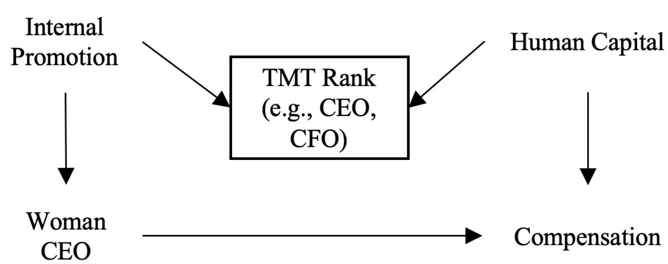


Figure 5. DAG of collider

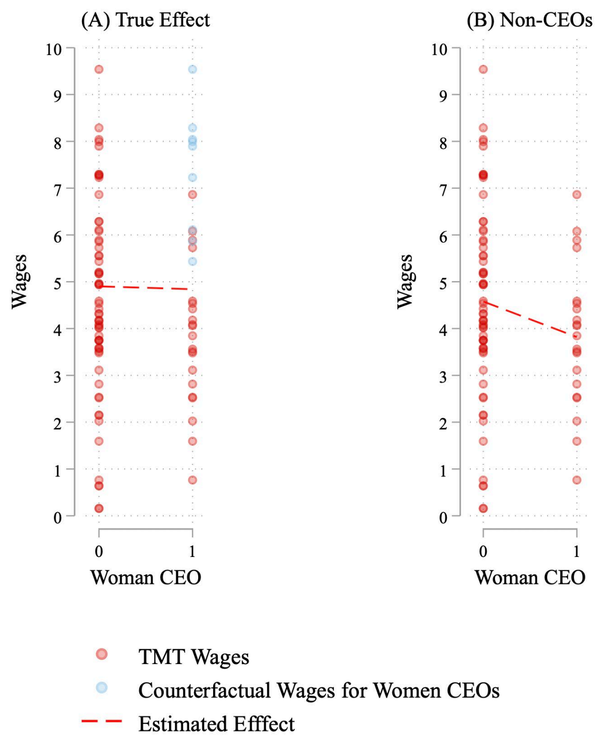


Figure 6: Illustration of Collider Bias Using Gender of CEO on the Compensation of non-CEO Women

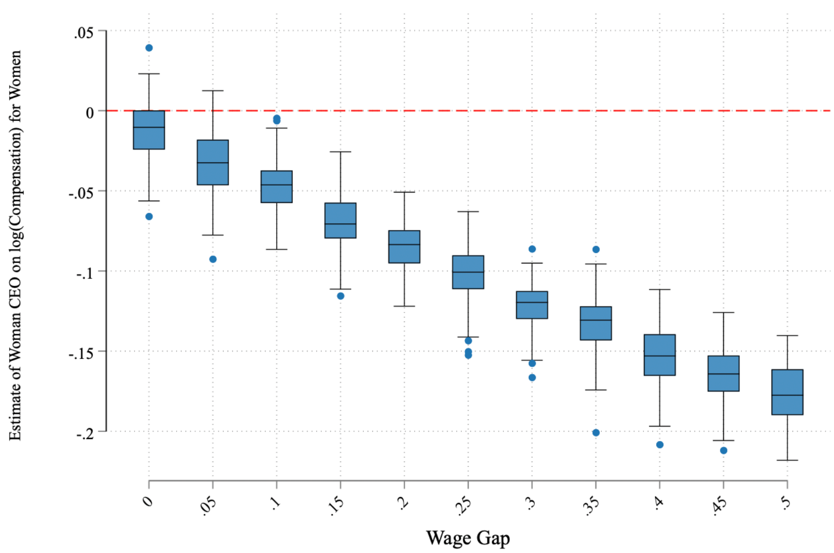


Figure 7: Estimated Effect of Women CEOs on the log(Compensation) of TMT Women, by Compensation Gap

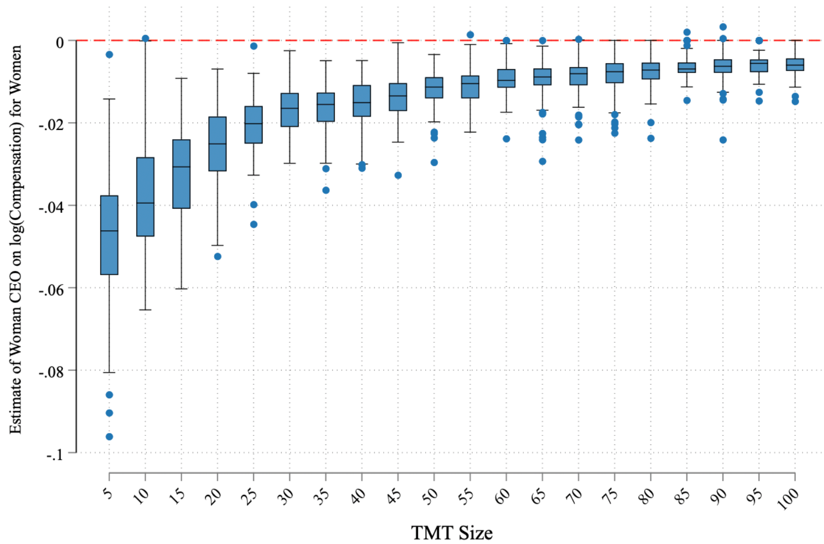


Figure 8: Estimated Effect of Women CEOs on the log(Compensation) of TMT Women, by TMT Size

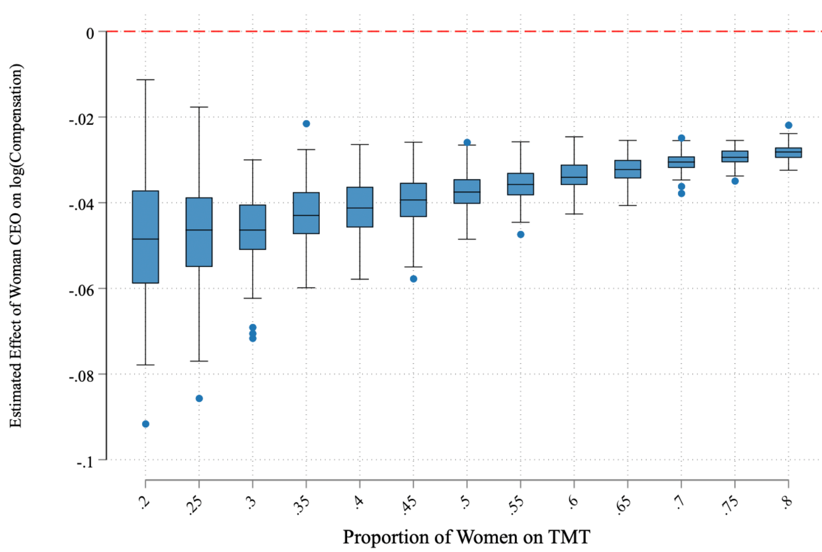


Figure 9: Estimated Effect of Women CEOs on the log(Compensation) of non-CEO Women, by Proportion of non-CEO Women on the TMT

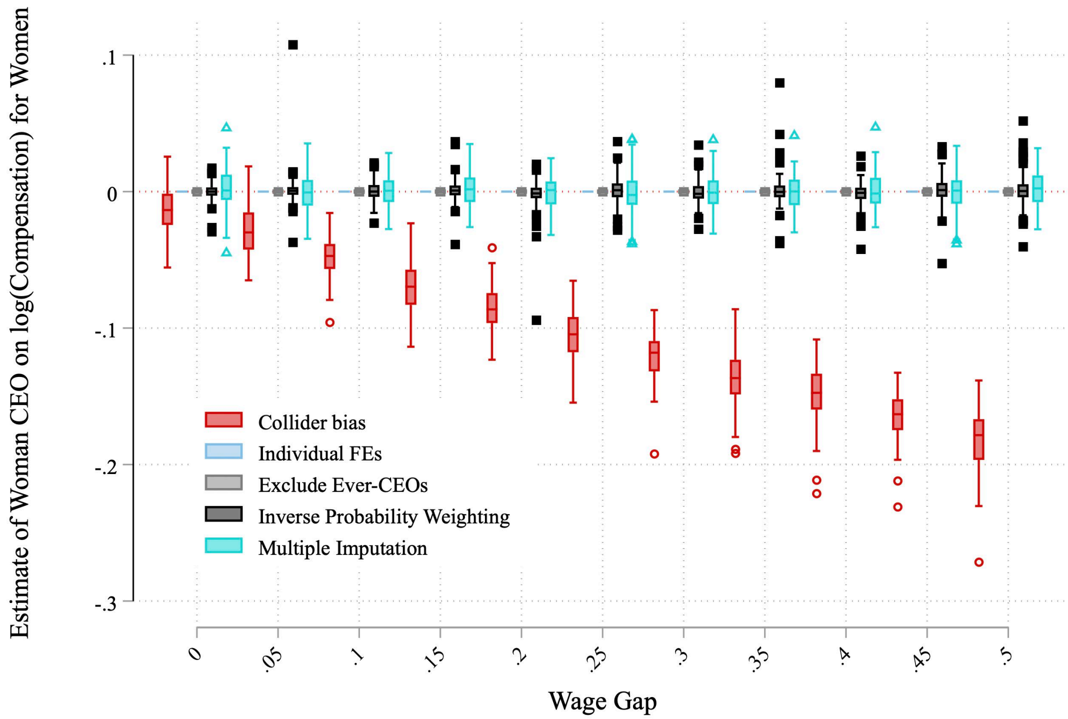


Figure 10: Estimated Effect of Women CEOs on the log(Compensation) of TMT Women, by Estimation Approach

Table 1. Frequency of Women, Minorities and Placebo Groups in Sample

	Observations	**Individuals**	**No. of firms**
Employee is a woman	11,335	2,686	1,171
Employee is minority	10,244	2,056	962
Employee first name is "John"	6,920	1,182	780
Employee name starts with "M"	14,033	2,482	1,181
Employee in random group	11,039	1,980	1,055

Table 2. Summary Statistics and Correlations (N = 143,236)

		Mean	SD	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
1.	Total Compensation	7.50	1.09															
2.	CEO is a woman	0.03	0.18	0.03														
3.	CEO in random group	0.07	0.26	0.03	-0.02													
4.	CEO's first name is John	0.05	0.21	0.02	-0.04	0.01												
5.	CEO's name starts with M	0.09	0.29	0.01	0.07	-0.03	-0.07											
6.	CEO is a minority	0.07	0.26	-0.01	0.01	0.01	-0.06	-0.01										
7.	Book Leverage	0.25	0.22	0.11	0.01	0.03	-0.01	-0.00	-0.03									
8.	Marketing Intensity	0.01	0.04	0.02	0.05	-0.01	-0.01	0.02	-0.00	0.01								
9.	R&D Intensity	0.03	0.06	-0.03	0.00	-0.03	-0.00	-0.01	0.06	-0.13	0.05							
10.	Size from assets	8.04	1.89	0.55	0.02	0.02	0.03	-0.01	-0.02	0.15	-0.12	-0.28						
11.	Age of Capital Stock	0.20	0.20	0.02	-0.02	-0.01	0.01	-0.01	0.05	-0.10	0.06	0.22	-0.17					
12.	Size from employees	2.13	1.40	0.46	0.01	0.00	0.02	-0.02	-0.03	0.08	0.05	-0.17	0.60	-0.07				
13.	Tobin’s Q	1.03	0.38	0.07	0.00	-0.00	0.01	0.01	0.02	-0.07	0.23	0.40	-0.29	0.16	-0.01			
14.	CapEx Intensity	0.05	0.06	-0.07	-0.01	0.01	-0.00	0.02	-0.04	-0.01	0.14	0.06	-0.16	-0.16	0.05	0.14		
15.	Executive's age	3.98	0.14	0.18	0.01	-0.00	0.00	-0.01	0.01	-0.00	-0.05	-0.06	0.14	-0.06	0.10	-0.05	-0.04	
16.	Employee is CFO	0.14	0.34	-0.02	0.02	0.00	-0.01	0.01	0.02	0.02	-0.01	-0.00	0.01	0.02	-0.04	-0.00	-0.05	-0.08

Notes:
These correlations correspond to the executive-firm-year level data used in the first analysis.
Correlations larger than |0.005| are statistically significant at $p < 0.05$.

Table 3. CEO Effect on the Compensation of Other Top Managers

Sample restricted to:	Other Women in TMT	Other minorities in TMT	Other "Johns" in TMT	Other "Starts with M" in TMT	Other "Random" in TMT
	Model 1	Model 2	Model 3	Model 4	Model 5
Book Leverage	-0.160	-0.168	-0.122	-0.143	-0.179
	(0.060)	(0.065)	(0.098)	(0.066)	(0.094)
Marketing Intensity	0.056	-1.024	0.301	-0.802	0.107
	(0.414)	(0.361)	(0.248)	(0.662)	(0.454)
R&D Intensity	-0.715	-0.325	-0.590	-0.092	-0.747
	(0.251)	(0.243)	(0.403)	(0.244)	(0.324)
Size from assets	0.149	0.131	0.098	0.162	0.115
	(0.028)	(0.029)	(0.041)	(0.027)	(0.020)
Age of Capital Stock	-0.103	0.091	0.036	0.006	0.030
	(0.056)	(0.096)	(0.102)	(0.069)	(0.104)
Size from employees	0.122	0.159	0.161	0.101	0.142
	(0.031)	(0.036)	(0.038)	(0.035)	(0.048)
Tobin's Q	0.222	0.297	0.396	0.317	0.269
	(0.034)	(0.057)	(0.075)	(0.043)	(0.053)
CapEx Intensity	-0.246	0.251	0.375	0.165	0.033
	(0.189)	(0.342)	(0.224)	(0.216)	(0.343)
Executive's age	0.376	0.297	0.695	0.608	0.453
	(0.095)	(0.155)	(0.159)	(0.111)	(0.164)
Employee is CFO	0.079	-0.002	0.073	0.066	0.033
	(0.028)	(0.036)	(0.041)	(0.027)	(0.041)
CEO is a woman	-0.171				
	(0.048)				
CEO is a minority		-0.285			
		(0.053)			
CEO's first name is John			-0.240		
			(0.069)		
CEO's name starts with M				-0.132	
				(0.040)	
CEO belongs to random group					-0.107
					(0.052)
Constant	3.322	3.802	2.126	2.271	3.246
	(0.470)	(0.622)	(0.680)	(0.474)	(0.669)
Observations	10257	8071	5383	11226	8855
Adj. R-squared	0.774	0.742	0.780	0.728	0.744

Notes:
Robust standard errors clustered by firm in parentheses.
All models include year and firm fixed effects.

Table 4. CEO Effect on TMT Representation

Dependent variable is:	% Women in TMT	% Minority in TMT	% "John" in TMT	% "Starts with M" in TMT	% Random in TMT
	Model 1	Model 2	Model 3	Model 4	Model 5
Book Leverage	0.021	0.006	0.002	0.001	0.011
	(0.018)	(0.017)	(0.008)	(0.011)	(0.010)
Marketing Intensity	0.074	-0.060	-0.016	0.163	-0.018
	(0.083)	(0.079)	(0.047)	(0.093)	(0.057)
R&D Intensity	0.049	0.019	-0.004	-0.056	-0.040
	(0.046)	(0.037)	(0.018)	(0.034)	(0.036)
Size from assets	-0.003	-0.003	-0.002	-0.002	-0.000
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Age of Capital Stock	0.013	0.011	0.004	-0.011	0.005
	(0.011)	(0.011)	(0.008)	(0.012)	(0.010)
Size from employees	-0.001	-0.002	0.001	-0.005	0.004
	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)
Tobin's Q	0.006	-0.006	0.002	-0.004	-0.004
	(0.007)	(0.006)	(0.004)	(0.007)	(0.006)
CapEx Intensity	-0.026	0.035	0.023	0.022	-0.011
	(0.022)	(0.029)	(0.022)	(0.024)	(0.024)
CEO is a woman	-0.037				
	(0.016)				
CEO is minority		-0.016			
		(0.014)			
CEO's first name is John			-0.016		
			(0.009)		
CEO's name starts with M				-0.041	
				(0.010)	
CEO belongs to random group					-0.030
					(0.008)
Constant	0.010	0.078	0.064	0.075	0.079
	(0.016)	(0.018)	(0.015)	(0.019)	(0.017)
Observations	26701	26701	26701	26701	26701
Adj. R-squared	0.438	0.513	0.348	0.381	0.374

Notes:
Robust standard errors clustered by firm in parentheses
All models include year and firm fixed effects.

Table 5. CEO Effect on Manager Compensation with and without Collider Bias

Dependent variable includes		Women	Minorites	First name is "John"	Name starts with "M"	In random group
	All TMT members (S.E.)	-0.171 (0.048)	-0.285 (0.053)	-0.240 (0.069)	-0.132 (0.040)	-0.107 (0.052)
	TMT members who never become CEO (S.E.)	-0.012 (0.043)	-0.063 (0.060)	0.035 (0.081)	0.005 (0.041)	0.081 (0.056)
	Percent obs. excluded (No. obs. excluded)	5.6% (577)	10.4% (836)	12.9% (697)	11.8% (1327)	12.3% (1088)

Table 6. The CEO Effect on TMT Representation with and without Collider Bias

Dependent variable includes:		Women	Minorites	First name is "John"	Name starts with "M"	In random group
	All TMT members (with collider bias)	-0.037 (0.016)	-0.016 (0.014)	-0.016 (0.009)	-0.041 (0.010)	-0.030 (0.008)
	TMT members who never become CEO	0.014 (0.014)	0.026 (0.012)	0.012 (0.007)	-0.001 (0.008)	-0.000 (0.007)