Collider Bias in Strategy and Management Research:
An Illustration of How Women CEOs Affect Other Women's' Career Outcomes

Justin Frake
Stephen M. Ross School of Business
University of Michigan
701 Tappan Street
Ann Arbor, MI 48109
jfrake@umich.edu

Andreas Hagemann
Stephen M. Ross School of Business
University of Michigan
701 Tappan Street
Ann Arbor, MI 48109
hagem@umich.edu

Jose Uribe*
Kelley School of Business
Indiana University
1309 E. 10th Street
Bloomington, IN 47405
uribej@iu.edu

*Equal contribution, authors listed in alphabetical order

ABSTRACT

Collider bias can cause spurious correlations when researchers condition on a variable that is caused by—or shares a common cause with—both the outcome and the exposure variable. We distinguish colliders from other threats to identification and estimation and illustrate its importance with a replication of published work suggesting that having a woman CEO reduces the career outcomes of other women executives. After accounting for collider bias, we find no evidence that women CEOs damage the career outcomes of other women in their organizations. We use Monte Carlo simulations to illustrate conditions that reduce or amplify the bias and provide generalizable approaches to identify and mitigate the risk of collider bias in applied 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. Colliders are variables that are caused -or share a common cause with- both the outcome and an independent variable. Collider bias occurs when the relationship between two variables is distorted by conditioning (e.g., controlling or selecting) on a collider variable (Elwert and Winship, 2014; Cinelli, Forney, and Pearl, 2022; Griffith et al., 2020; Schneider, 2020). A number of published papers in economics have been criticized for the potential presence of collider bias, including work on the influence of plantations and the slave trade on settler mortality rates (Acemoglu, Johnson, and Robinson, 2001), racial bias in policing (Fryer, 2019), and the literature on the "early industrial growth puzzle" (Kolmos, 1998). Despite the potential impact of collider bias on the validity of empirical results in strategy and management research, its effect is often overlooked.

We formally define collider bias and explain differences from biases induced by confounders, mediators, and other identification and estimation issues. Collider bias is distinct from sample selection bias (Shaver, 1998; Certo et al., 2016) caused by a confounder, or unobserved variable affecting both the treatment and the outcome. By contrast, collider bias occurs when selecting (or otherwise conditioning) on a variable that is itself influenced by—or shares a common cause with—both the outcome and the treatment. Collider bias is also distinct from other estimation issues, like multicollinearity (e.g., Kalnins 2018), which is caused by a too-close relationship between two or more independent variables that distorts estimates of their effect on the outcome.

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 woman or minority CEO 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).

We carry out an empirical demonstration of collider bias using the well-known ExecuComp and Compusat datasets to examine the relationship between women CEOs on both the prevalence and compensation of women 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 in the top management team.[1]

We argue that the effects we find are a result of conditioning on a collider, namely the propensity of women to exit the sample when promoted from non-CEO to CEO positions. Rather than gender dynamics, the data-generating process includes unexpected panel attrition: promotion into the CEO position implies exit from the sample. To support this claim, we show

[1] For the remainder of this article we will refer to Dezso, Li and Ross (2022) as DLR and to Corwin, Loncarich and Ridge (2022) as CLR.

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. Moreover, we answer Shaver's (2020) request to go beyond calls for replication, showing that replication efforts that do not attend to the specific source of collider bias will yield robust coefficients of misleading results.

Through Monte Carlo simulations, we show that collider bias can generate significant spurious correlations. We also demonstrate that the extent of this bias depends on three factors: the proportion of women in the TMT, the TMT's size, and the compensation differential between individuals who will become CEOs in future periods and other TMT members. We advocate the use of Directed Acyclic Graphs (DAGs) for identifying potential sources of collider bias. We also assess various corrective methods in this setting, including individual fixed effects, the exclusion of individuals destined for future promotions, inverse probability weighting, multiple imputation, and placebo analysis. Our discussion further delves into how collider bias intersects with other identification challenges such as bad controls, selection on the dependent variable, nonresponse bias, and attrition bias.

Our study draws attention to the importance of considering collider bias in empirical research in management and strategy. 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 into the CEO position. Broader awareness can therefore increase the validity and usefulness of our findings for informing theory and practice.

Perhaps the most fundamental aim is a renewed call for empiricists to attend to identification and the nature of the data-generating process (see also Shaver, 2020). While other types of empirical problems such as classical measurement error often merely attenuate effects, collider bias can reverse the sign on estimated coefficients. The widely acknowledged publication pressures on faculty, particularly pre-tenured faculty, and the field's penchant for surprising and counterintuitive findings (e.g. Davis, 2015) makes attention to collider bias particularly timely.

2 | BACKGROUND

The primary aim of most empirical studies in strategy and management is to accurately estimate unbiased causal effects (Shaver 1998, 2020, 2021; Hill et al. 2021). A stream of work has underscored the necessity of accounting for sources of bias (Shaver 1998, Hamilton and Nickerson 2003; Stern et al, 2021) such as omitted variable bias (Busenbark et al. 2022, Wolfolds and Siegel, 2019), selection bias (Certo et al. 2016), measurement error (Boyde, Grove, and Hitt, 2005; Ge et al. 2016), and multicollinearity (Kalnins 2018). Despite the growing attention to estimation and identification, the issue of collider bias remains a relatively unexplored threat to causal inference. To illustrate this point, we compare attention to confounders and attention to colliders in articles published between 2010 and 2023 in key journals across the fields of Strategy, Economics, and Sociology. In each of the selected top journals, we conducted a search for mentions of potential bias from confounders and from colliders. The results of this exercise are presented in Table 1.

[INSERT TABLE 1 HERE]

There is a relative under-representation of terms associated with collider bias in prominent Management and Strategy journals, compared to prominent journals in Economics and Sociology. Only 0.68% of articles in Management and Strategy journals mention collider-related terms, a rate that is about 50% lower than Economics journals and 80% lower than Sociology journals. When it comes to confounders, 19% of Strategy and Management articles mention the term, which is only 40% lower than Sociology and 50% higher than Economics. These comparisons are noteworthy for two reasons. First, the types of econometric models and data used in Strategy and Management are just as likely to suffer from collider bias as those used in Economics and Sociology. Second, threats to causal inference from collider bias are as important as the well-known biases that arise from omitting confounders. We next illustrate the distinction between confounders and colliders using directed acyclic graphs, an approach originally developed by Pearl (2000).[2]

2.1 | Using DAGs to understand collider bias

Directed Acyclic Graphs (DAGs) are a visual tool for mapping the causal connections among variables. DAGs rely on simple graphical rules: each node symbolizes a random variable, while arrows indicate direct causal connections. An absence of arrows between two nodes signifies that no causal link exists between the variables they represent. The term "path" denotes an ordered sequence of arrows that connect two variables.

When all arrows on a path point from the treatment variable to the outcome variable, it is referred to as a "causal path." In general, causal identification aims to identify the total effects of the treatment across all causal paths. For example, in Figure 1A the total causal effect of the treatment (X) on the outcome (Y) would include the following causal paths: X → Y and X → M → Y. Paths that are not causal are known as "backdoor paths," which can induce bias when

[2] Online Appendix A contains a more detailed discussion of Directed Acyclic Graphs (DAGs).

estimating the relationship between the treatment and the outcome. For example, X ← U → Y (in Figure 1B) and X → C ← Y (in Figure 1C), are backdoor paths.

[INSERT FIGURE 1 HERE]

Besides the treatment and outcome variables, three basic types of variables exist in a DAG: mediators, confounders, and colliders. Mediators are variables found in a causal path. In Figure 1A, M is a mediator because it is influenced by X and subsequently affects Y. Confounders exist on backdoor paths and cause two or more other variables. In Figure 1B, U is a confounder because it causes X and Y. Colliders also exist on backdoor paths but are influenced by two or more other variables. In Figure 1C, C is a collider because it is caused by X and Y.

In the context of a DAG, the objective of causal identification is to ensure that all causal paths remain open, allowing for unobstructed analysis of direct relationships between variables, while closing all backdoor paths that could introduce bias. Conditioning—controlling, selecting, or stratifying on a variable—is the primary method used for opening and closing paths. The effect of conditioning varies depending on the type of variable involved. Conditioning on mediators and confounders closes the path. For instance, conditioning on M in Figure 1A would lead to overcontrol bias by closing the causal path X → M → Y. Further, failing to condition on U would induce omitted variable bias, sometimes also referred to as sample selection bias. Conditioning on colliders opens the path. In Figure 1C, the path X → C ← Y is already closed when C is not conditioned upon because C absorbs the variation from X and Y. However, conditioning on C opens this backdoor path and allows for a spurious correlation between X and Y, leading to collider bias, which is the focus of this paper.

2.2 | Illustrative example

To make the concept of collider bias more concrete, consider the example of studying the effect of Corporate Social Responsibility (CSR) on a firm's financial performance. Despite the existence of over 2,200 empirical papers on this topic (Friede, Busch, and Bassen, 2015), scholarly opinion remains divided (Awaysheh et al., 2020). While some propose a negative relationship, contending that CSR benefits stakeholders at the expense of shareholders (Friedman, 1970), others posit a positive effect due to the reputational advantages of CSR Hornstein and Zhao, 2018). Scholars like McWilliams and Siegel (2000) attribute this lack of consensus to confounders and omitted variable bias, whereas Awaysheh et al. (2020) point to measurement issues. However, collider bias may be just as severe a threat in many studies.

To illustrate this, imagine a researcher using the MSCI (KLD) database to obtain CSR metrics to construct their key independent variable.[3] This researcher also acquires a complete and reliable dataset to accurately measure revenues for all firms. After carefully controlling for all confounding variables such as industry and R&D (McWilliams and Siegel 2000), the researcher runs a regression and estimates a significant, positive effect of CSR on revenues. Because not all firms have CSR metrics in the MSCI (KLD) database, authors will typically acknowledge the sample's limitations with a disclaimer such as "Among firms that voluntarily disclose their CSR activities, CSR increases firm performance." In other words, after acknowledging possible threats to generalizability, the estimate is deemed valid for the sample at hand. However, this conclusion may be wrong due to collider bias.

Collider bias in this case stems from sample selection based on firms' CSR disclosures. This is not a minor concern affecting merely the generalizability of the results; rather, it renders the observed CSR effect on financial performance spurious even within the selected sample. For

[3] Other similar datasets include Refinitiv ASSET4, Dow Jones Sustainability Index, and Sustainalytics.

further clarity, let us assume that the true causal relationship between CSR and financial performance is negative (as shown in Panel A of Figure 2). The MSCI (KLD) database relies on voluntary CSR disclosures via various channels, including company websites and annual reports. With that in mind, suppose that firms with strong CSR records are more likely to disclose their CSR activities, and thus be over-represented in the MSCI (KLD) database. Suppose also that firms with weak financial performance are more likely to emphasize their CSR activities as a compensatory strategy. Given these assumptions, the estimated positive effect of CSR on financial performance among the disclosing firms (as seen in Panel B of Figure 2) is spurious. It is not a genuine causal effect but rather a consequence of conditioning on a collider—CSR disclosure. This relationship can be graphically illustrated in the DAG in Figure 1C by letting X, Y and C respectively represent CSR activities, financial performance, and CSR disclosure.

[INSERT FIGURE 2 HERE]

2.3 | Collider bias in published work

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., DLR). 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 a 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 3A, 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 wrongly conclude that having a woman CEO increased the compensation of women TMT members by $25,000.

[INSERT FIGURE 3 HERE]

To address this problem, DLR exclude CEOs from their sample and condition the analysis on non-CEO women only, as seen in figure 3B. 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 compensation.

We represent a simplified DAG of the assumed causal structure in figure 4A (online appendix A contains the full DAG including all steps in its construction). *Internal promotion* of a woman in the TMT causes the TMT member's rank to change (e.g., from CFO to CEO) and also causes there to be a woman CEO. There is also a direct effect of *TMT rank* on *compensation*. The variable *Internal promotion* refers to the unobserved opportunity structure at a firm by which a focal woman may come to occupy the CEO position, such as the previous CEOs retirement or firing, and the availability of other suitable candidates. 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 exists along two backdoor paths. The first is the path highlighted in figure 4B, where *internal promotion* is a confounder and *TMT Rank* is a mediator. By conditioning on the mediator *TMT rank*, the authors effectively close the backdoor path highlighted in figure 4B. However, in doing so, they open another path, highlighted in figure 4C. Specifically, because *TMT rank* is a collider between internal promotion and human capital, and because conditioning on a collider opens a path, conditioning on *TMT rank* opens the backdoor path highlighted in figure 4B. Thus, by conditioning on *TMT rank*, the researchers have closed one biasing path, but inadvertently opened another. Note that in this particular causal model, *TMT rank* is both a collider and a confounder. This highlights the importance for researchers to explicitly state their assumed causal model.

[INSERT FIGURE 4 HERE]

Next, we generalize and formalize the intuition in this example by deriving the bias term using a simple mathematical model. As in figure 3, 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 premium 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 compensation:

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

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

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

where we used $W = Y_{2,1} - Y_{1,1}$ and $\delta = E(W) > 0$ as defined above (1). 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. The bias persists even though the fixed effects estimator successfully eliminated the confounder V. 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 $Y_{1,t}, \ldots, Y_{n-1,t}$ 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 5, 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 (different) woman CEO. 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 5A. However, figure 5B 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 5 HERE]

3 | EMPIRICAL ILLUSTRATION

Until now, we have used thought experiments, simulated data, DAGs, and mathematical models to illustrate the problem of collider bias 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 DLR, who find that the presence of a woman CEO (compared to a man CEO) reduces the compensation of other women on the TMT by more than 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, a different set of authors (CLR) 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 gender diversity of their TMT 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 DLR and CLR.

While in line with the findings in extant work, we aim to demonstrate that the observed correlations we report are unrelated to the gender of the CEO and can be explained by collider bias. Again, CEO rank 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 produce comparable effects on the outcomes of non-CEOs sharing the same characteristic. Later we will 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[4]

[4] In online appendix B we replicate the main results from DLR and CLR across various specifications, and explain differences in sampling and estimation. In online appendix D we provide sample Stata code to replicate the main DLR results.

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. DLR) and 2) the proportion of other women in the TMT (cf., CLR). We select all firms and employees with valid compensation data in the ExecuComp dataset, which collects information on the highest paid employees in S&P 1500 firms and various other firms. We construct the TMT as consisting of the CEO and any additional employees reported, who typically hold titles such as COO, CFO, Executive VP, etc. Our final sample consists of 298,975 observations for 55,077 executives working at 3,960 firms. The ExecuComp dataset includes a variable indicating whether an executive is a man or a woman. Observations for women employees constitute 7.40% of the total sample.

3.2 | Placebo groups

Placebo groups are unrelated to the gender dynamics advanced by DLR and CLR. Thus, if placebo regressions exhibit similar correlations to those found for women, 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 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" (5.00% of the total sample), employees whose first name starts with the letter "M" (9.00% of the sample), and a group of randomly selected employees (7.9% of the sample).[5]

Women in our sample are largely non-overlapping with the placebo groups: The "John" group contains 0% women observations, the "Letter M" group contains 6.5% women

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

observations, and the “Random” group contains 6.9% women observations. Table 2 below describes the size of the various groups used in the analyses.

[INSERT TABLE 2 HERE]

3.3 | Variable definitions

3.3.1 | Manager compensation and TMT representation

We followed DLR in constructing *Top Manager compensation* (Dependent Variable 1) as the natural log transformation of a top manager’s total compensation, including salary, bonus, and grants of stock and options. We followed CLR in constructing *% Women in TMT* (Dependent Variable 2) 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 *% “John” in TMT, % “Letter M” in TMT* and *% Random Group in TMT.*

3.3.2 | Independent variable of interest: CEO types

Has CEO of Type X. We directly followed DLR and CLR in the construction of the independent variable *Woman CEO,* which takes the value 1 if, in a given year, a firm has a woman CEO and 0 otherwise. We construct parallel measures for *John CEO, Letter M CEO* and *Random Group CEO.*

3.3.3 | Control variables

We followed DLR in the selection and measurement of all firm-level covariates and two manager characteristics as explained below.

Advertising intensity is the log transformation of the ratio of advertising expense to assets. *Firm age* is the log transformation of the firm’s age in years, measured as the difference between the current year and the earlier of the firm’s first year in CompuStat or initial public trading date.

Leverage is the ratio of debt to the market value of a firm's assets. *R&D intensity* is the log transformation of the ratio of R&D expense to assets. *Size from assets* is the log transformation of the lagged book value of a firm's assets. *Size from employees* is the log transformation of the lagged size of a firm's workforce. *Tobin's q* is the log transformation of the lagged ratio of the market value of a firm's assets to their replacement value. *Manager age* is the log transformation of the manager's age. *Employee is CFO* is a dummy variable indicating that the employee has the title of Chief Financial Officer, a variable which is readily available in the ExecuComp dataset. Table 3 below contains descriptive statistics and zero order correlations for the variables in our analyses.

[INSERT TABLE 3 HERE]

3.4 | Results

3.4.1 | The Effect of Women 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 CEO on the total compensation of other women on the TMT. As in DLR, we run the analysis in the unpooled sample corresponding to the group sharing the same characteristic as the CEO.[6]

Model 1 in table 4 shows a negative coefficient of -0.131, which corresponds to a 12.3% decrease in the compensation of other women in the TMT from having a woman CEO compared to a man CEO. This effect is similar in size to the estimated effect reported by DLR (cf. table 5, model 2 in DLR). These regressions do not, however, demonstrate that collider bias is driving the effects. It is possible that women CEOs do have real negative effects on other women TMT members. To explore whether this is the case, we turn to placebo regressions in models 2-4.

[6] DLR also show a null effect for having a woman CEO on the compensation of men. These results are unproblematic with respect to collider bias. This is because in the ExecuComp dataset non-CEO men do not exit the sample to become women CEOs.

The results are quite similar for the placebo groups. Individuals named “John” (model 3) appear to suffer an 18% 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” (-10%) or who belong to a random group (-15%), from having a CEO sharing that specific trait. Unless we are to believe that there are causal mechanisms causing CEOs named John to reduce the compensation of other TMT members named John, the results cast doubt on the causal explanation that women CEOs hurt the compensation of other women executives.

3.4.2 | The Effect of Woman CEOs on TMT Gender Diversity

Next, we follow CLR and collapse the data into firm-year observations to analyze the effect of having a woman CEO on the proportion of other women on the TMT in the subsequent year. Table 5 displays the coefficients for the analyses on women and on the three placebo groups.

[INSERT TABLES 4 AND 5 HERE]

The magnitude of the effect of a woman CEO on the proportion of women on the TMT in the subsequent period is -3.1%, about half of the negative effect reported by CLR. The effect of women CEOs on gender representation is very similar to the effect of CEOs whose name starts with the letter M on the proportion of employees whose name also starts with the letter M (-3.8%, see model 2 in table 5). The effect of a CEO named “John” and of a CEO from a randomly assigned group is also negative and precisely estimated on the proportion of employees named “John” (-2.0%), and on employees from the same randomly selected group (-2.0%). 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 CEOs erode the gender diversity of top management teams.

3.5 | Correcting collider bias

3.5.1 | Excluding Individuals Who Will Become CEOs in Future Periods

The solutions to collider bias vary by context. The simplest solution is to avoid conditioning on colliders. If we return to the simplified DAG in figure 4, it is possible to develop an identification strategy. As we explained before, conditioning on TMT rank opens a backdoor path because TMT rank is a collider. However, conditioning on internal promotions closes the backdoor path identified in figure 4B without opening the backdoor path highlighted in figure 4C. Thus, the DAG makes it clear that the best solution is to condition on *internal promotion*. For our empirical example, this can be done by excluding (selecting out) all women who were *internally promoted to CEO*. Another effective approach is to condition on TMT rank *and* include individual-level fixed effects, which effectively closes the backdoor path highlighted in figure 4C by conditioning on the TMT member's human capital (assuming human capital is relatively stable over the sample period). As shown in online appendix C, individual level fixed effects greatly reduce the contribution of women who are internally promoted to CEO to the estimated coefficient. In tables 6 and 7 we compare the coefficients and p-values previously reported in tables 4 and 5 with those of identical regressions run on samples that exclude individuals who will become CEOs in future periods.

Table 6 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 and the three placebo groups all become very small and statistically indistinguishable from zero.

[INSERT TABLE 6 HERE]

Similarly, table 7 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 statistically indistinguishable from zero across all groups.

[INSERT TABLE 7 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 -0.198 to -0.110 in Table 6 and from -0.038 to -0.020 in table 7. Next, we use Monte Carlo analysis to explore the conditions under which the bias is more or less severe.

4 | SIMULTION AND 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 individuals who will become CEOs in future periods experience prior to being promoted to CEO.

4.1 | Simulation overview

The baseline simulation follows the data generating process in figure 4, with the exception that we assume no effect of *woman CEO* on *women's compensation*. Thus, the "true" causal effect in this simulation is always zero. Next, we generate a binary variable woman to indicate whether an individual is a woman, with a 10 percent probability. Each individual is assigned a unique ID and exists in the data for two periods. A firm ID is generated such that each firm has five employees over two periods. A human capital variable is generated, ranging from 0 to 5, and is kept constant across the two periods for each individual. Individuals are ranked within firms based on their human capital; the person with the highest human capital gets the highest rank. In the second period, the highest-ranked individual is promoted to CEO. Lastly, a compensation variable is generated as a function of *human capital*, *rank*, and a random error term.[7]

To show that this simulation works as intended, we simulate a large dataset and run several regressions to demonstrate how various approaches to selection and statistical control

[7] The data-generating process is described by the following equations:

- $\text{Woman}_{i,j} \sim \text{Bernoulli}(0.2)$
- $\text{Period} \in \{1,2\}$
- $\text{HumanCapital}_{i,j} \sim \text{Uniform}(0,5)$
- $\text{Rank}_{2,j} = f(\text{HumanCapital}_{\text{i}}, \text{Rank}_{1,\text{j}})$
- $\text{Compensation}_{i,j} = \text{HumanCapital}_{i,j} + \text{Rank}_{i,j} + \epsilon_{i,j}$ with $\epsilon_{i,j} \sim \text{Normal}(5,1)$

affect the accuracy of the estimation. Table 8 presents regression results examining the effect of having a woman as a CEO on log(compensation) of women TMT members.

[INSERT TABLE 8 HERE]

Model 1 shows the effect of a Woman CEO on log(compensation) without adjusting for rank or human capital. This model suggests that firms led by women CEOs pay women higher compensation. However, this estimate is biased because both backdoor paths identified in Figures 4B and 4C are still open. Model 2 adjusts for rank by excluding CEOs from the sample, as DLR did. The regression finds a precisely estimated negative effect of having a Woman CEO, in line with their findings. Model 3 adjusts for human capital using individual fixed effects, but does not adjust for rank and returns a precisely estimated positive effect.

Models 4 both excludes CEOs and includes individual fixed effects, thereby closing both biasing paths identified in figure 4. Model 5 excludes women who were ever CEO. As expected, both models 4 and 5 return small coefficients that are indistinguishable from zero, suggesting that both approaches are sufficient to close the biasing paths identified in figure 4.

Next, we use this simulation to run Monte Carlo analysis. For each scenario, we simulate 100 datasets with 1,000 observations each according to the data-generating process described above. Again, the "true" causal effect of a woman CEO on others' compensation as defined by the data generating process is zero. For the baseline simulations, we also assume that the compensation premium that CEOs receive is 10 percent, each TMT consists of five people,[8] and that 10 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

[8] We do this following DLR, who use the top five most highly paid TMT members in their analysis.

In figure 6, we vary the compensation premium that individuals who will become CEOs in future periods 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 bias increases as TMT members with the wage premium paid to people who will become CEOs.

[INSERT FIGURE 6 HERE]

Having shown how differences in the compensation premium (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 7, we set the compensation premium 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 7 HERE]

In figure 8, we again fix the compensation premium at 10 percent and the number of non-CEO TMT members at five. We vary the proportion of women in top management teams from 10 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 8 HERE]

5 | PRACTICAL ADVICE FOR AVOIDING COLLIDER BIAS

5.1 | Using Directed Acyclic Graphs (DAGs)

Because of the fundamental problem of causal inference—that we can never observe what would have happened to the treated unit had they not been treated—researchers can never know whether their specifications are biased. However, a well-articulated causal model can lead to a specification that delivers unbiased estimates, assuming the model itself is accurate. To clearly represent and articulate this assumed causal model, we advocate that researcher map their assumed causal model using DAGs. DAGs are particularly beneficial in observational studies, where the lack of experimental control can complicate causal inference. DAGs also help researchers identify the minimal set of variables that must be controlled for to obtain an unbiased estimate of a causal relationship between a particular independent and dependent variable (assuming the DAG is accurate). Moreover, DAGs offer a systematic method for sensitivity analysis, allowing researchers to explore alternative causal pathways and evaluate the robustness of their findings. For instance, if an omitted variable is a concern, adding it to the DAG helps researchers assess its potential impact on the causal estimates. Finally, the explicit articulation of the causal model that the researcher has in mind enhances the transparency, rigor, and credibility of empirical research.

While a comprehensive discussion of DAGs is outside the scope of this paper, we refer interested readers to Chapters 6-9 of Huntington-Klein (2022) and Hünermund et al. (2023). For a hands-on introduction, we also include a DAG primer in online appendix A, which describes how we developed the more complete DAG underlying figure 4. Briefly, researchers should start by identifying the treatment, outcome, and other relevant variables that either cause the treatment and outcome variables. Then, they should connect these nodes with directed arrows to signify the causal links, ensuring that the arrows flow from cause to effect and do not form feedback loops.

After constructing the DAG, they should examine it to identify colliders, which are variables affected by two or more other variables, identifiable by incoming arrows from multiple sources.

5.2 Checking typical hiding spots for collider bias

Collider bias manifests in various forms and intersects with a range of empirical issues, including but not limited to bad controls, selection bias, selection on the dependent variable, nonresponse bias, and attrition bias. Table 9 summarizes the most likely forms of collider bias in strategy and management research.

[INSERT TABLE 9 HERE]

Controls, matching, and fixed effects. When conditioning on a collider through statistical controls, including fixed effects, researchers essentially introduce what is known as a "bad control" (Angrist and Pischke, 2009; Cinelli, Forney, and Pearl, 2022). For instance, when examining the effect of CSR on performance, one should avoid controlling for a firm's inclusion in Forbes' Most Admired Companies list, as both CSR and performance likely influence a firm's likelihood of being featured on the list. Similarly, techniques like propensity score matching or coarsened exact matching can induce collider bias if the researcher matches on a collider. Our general recommendation is to only use controls, fixed effects, or matching variables that occur and are measured prior to treatment. Any control variables that occur after treatment are likely to be mediators or colliders, both of which researchers should generally avoid conditioning on.

Selection into datasets. Collider bias can be particularly difficult to detect when it arises through the sample selection process. The most egregious example is selection on the dependent variable. While most researchers acknowledge that selecting on the dependent variable can introduce bias in estimates, fewer recognize this as a special instance of collider bias. In such cases, the dependent variable serves as a collider between the treatment variable and the error term. By conditioning on the dependent variable, researchers open the backdoor path from the independent variable to the error term, thereby inducing bias. For example, when investigating the impact of firm size on employee turnover, researchers should refrain from using data sourced from public LinkedIn profiles (e.g., Revelio) because people likely select into making their profiles public around the time they intend to leave their employer.

Collider bias can also result from selecting on other variables that are caused by the dependent variable. As elaborated in Section 2.2, selection into the MSCI (KLD) database acts as a collider because a firm decides whether to disclose its CRS based on its CSR activity and its financial performance. As another example, Miller, Le Breton-Miller, and Lester (2010) use data from the largest 1000 firms in the U.S. (Fortune 1,000) to find that family ownership is negatively related to acquisitions. Unless firm size is causally independent from acquisition intentions or from family ownership, selecting on large firms may induce collider bias. For example, if family firms are likely to be smaller and if acquisition make firms larger, then conditioning on the largest firms acts as a collider. This can result in a negative correlation between family ownership and acquisitions even if the actual correlation in the universe of all firms is positive or nonexistent. To mitigate collider bias from selection into a dataset, researchers should be explicit about the dataset's inclusion criteria. After articulating a complete DAG, researchers can carefully consider whether the inclusion criteria might act as a collider in the data-generating process.

Sample exclusion criteria. Scholars should also be careful to avoid selecting on colliders when excluding observations from their sample. For instance, strategy and management scholars often drop observations for which they do not have control variables. If the reason for not having control variables is related to a collider, then the decision to exclude observations could result in collider bias. For instance, several studies explore the link between an inventor's patenting history and their likelihood to change employers (Hoisl, 2007; Palomeras and Melero, 2010; Melero, Palomeras, and Wehrheim, 2020). These studies often limit their samples to inventors who have patented at least twice. This constraint is due to the need for a minimum of two patents to measure a single mobility event—one to identify the originating firm and another to identify the destination firm. However, if the first patent increases the likelihood of subsequent patenting, and if changing employers also influences the propensity to patent (Kaiser, Kongsted, and Rønde, 2015), then this sample restriction based on the total number of career patents can introduce collider bias.

Attrition and nonresponse bias can also operate as a form of collider bias when leaving the sample or participating in a survey is influenced by both the independent and dependent variables. Imagine a study where a researcher surveys CEOs to examine whether CSR activities lead to improved financial performance. If CEOs who are more altruistic are both more likely to engage in CSR and more inclined to respond to the survey, and if hard working/busy CEOs are more profitable and are less likely to respond to surveys, then the study could be compromised by collider bias arising from nonresponse.

5.3 | Approaches to addressing collider bias

In the following section we present a number of tools that may assist researchers in ensuring the integrity of their empirical results. While 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.

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., 2022). 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 9, 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 for 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 4). 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 to correct for their probability of being promoted. As demonstrated in Figure 9, 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 compensation of women in the years that they were CEOs based on their human capital. As seen in Figure 9, 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 9 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.

6 | 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 non-CEOs from having a CEO who is a woman 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 on the compensation and representation of other women in the top management teams of firms in the ExecuComp dataset. Our results for women CEOs are similar in terms of magnitude and statistical significance to recent work. We provide evidence that collider bias – rather than the gender of individuals – drives our results. Without awareness of the collider bias problem, we would wrongly conclude that a woman CEO reduces the compensation and prevalence of other women in the TMT.

Recent replication efforts in strategic management have tested and redefined the robustness and scope conditions of previously published claims (cf. Bettis et al., 2016). Replication efforts in strategic management typically end up restricting a claim's scope conditions by extending the sampling and analytical strategy in various directions, including a longer sampling period (e.g. Howard et al, 2016), a broader population (e.g. Kalnins, 2016) or more up-to-date model specifications (e.g., Park et al, 2016). Attention to collider bias can complement these efforts by uncovering potentially misleading findings that are surprisingly robust to statistical replication. In online appendix B, we replicate the main result from DLR and CLR after extending the sample to include additional years of data, a larger pool of executives, and after removing all time-varying controls. In the main paper we also replicate the main pattern of results even after changing the focal population from women executives to executives named John, executives whose last name starts with the letter "M" and a group of randomly selected executives.

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 for individuals who will become CEOs in future periods 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 a woman 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 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.

Bettis, R. A., Helfat, C. E., & Shaver, J. M. (2016). The necessity, logic, and forms of replication. Strategic Management Journal, 37(11), 2193-2203.

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.

Boyd, B.K., Gove, S. and Hitt, M.A., 2005. Consequences of measurement problems in strategic management research: the case of Amihud and Lev. Strategic Management Journal, 26(4), pp.367-375.

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

Busenbark, J.R., Yoon, H., Gamache, D.L. and Withers, M.C (2022). Omitted variable bias: Examining management research with the impact threshold of a confounding variable (ITCV). Journal of Management, 48(1), pp.17-48.

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. (2022). 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.

Davis, G. F. (2015). Editorial essay: What is organizational research for?. Administrative Science Quarterly, 60(2), 179-188.

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.

Hill, A.D., Johnson, S.G., Greco, L.M., O'Boyle, E.H. and Walter, S.L., (2021). Endogeneity: A review and agenda for the methodology-practice divide affecting micro and macro research. Journal of Management, 47(1), pp.105-143.

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

Hornstein, A. S., & Zhao, M. (2018). Reaching through the fog: Institutional environment and crossborder giving of corporate foundations. Strategic Management Journal, 39(10), 2666–2690.

Howard, M. D., Withers, M. C., Carnes, C. M., & Hillman, A. J. (2016). Friends or strangers? It all depends on context: A replication and extension of Beckman, Haunschild, and Phillips (2004). Strategic management journal, 37(11), 2222-2234.

Huenermund, P., Louw, B., & Rönkkö, M. (2022). The Choice of Control Variables: How Causal Graphs Can Inform the Decision. In Academy of Management Proceedings (Vol. 2022, No. 1, p. 15534). Briarcliff Manor, NY 10510: Academy of Management.

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.

Kalnins, A. (2016). Beyond Manhattan: Localized competition and organizational failure in urban hotel markets throughout the United States, 2000–2014. Strategic Management Journal, 37(11), 2235-2253.

Kalnins, A. (2018). Multicollinearity: How common factors cause Type 1 errors in multivariate regression. Strategic Management Journal, 39(8), 2362-2385.

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.

Park, U. D., Borah, A., & Kotha, S. (2016). Signaling revisited: The use of signals in the market for IPO s. Strategic Management Journal, 37(11), 2362-2377.

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.

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.

Shaver, J. M. (2020). Causal identification through a cumulative body of research in the study of strategy and organizations. Journal of Management, 46(7), 1244–1256.

Shaver, J. M. (2021). Evolution of Quantitative Research Methods in Strategic Management. In I. M. Duhaime, M. A. Hitt, & M. A. Lyles (Eds.), Strategic management: State of the field and its future (pp. 83–97). Oxford University Press.

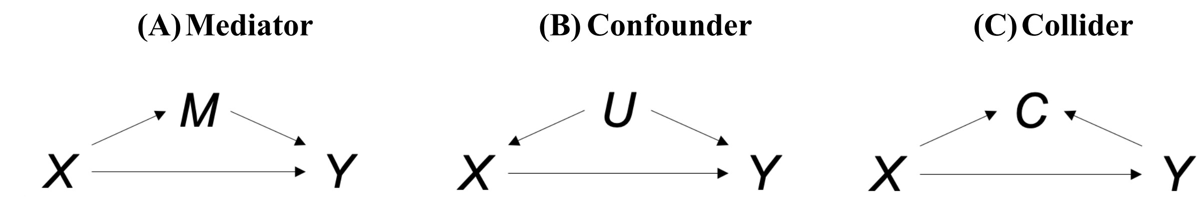
Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. Royal Society open science, 3(9), 160384.

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.



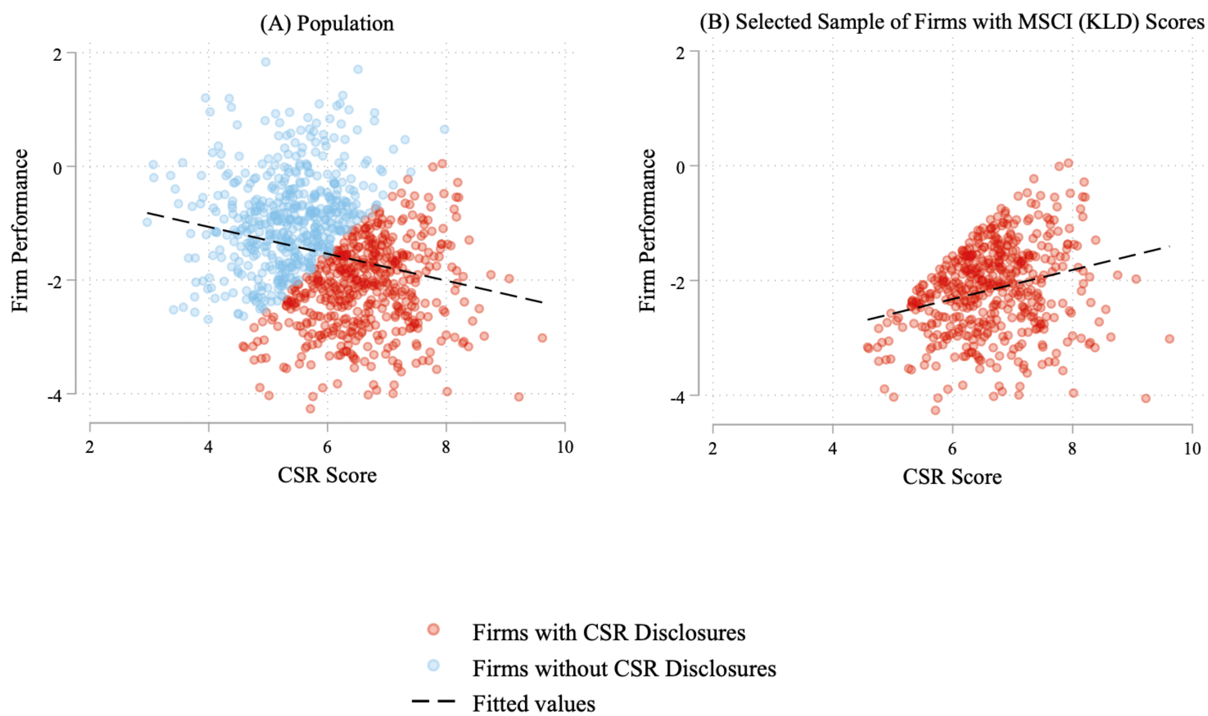
M is a mediator. Conditioning on M causes overcontrol bias.

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

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

Figure 1. Illustrating Confounders and Colliders using DAGs[9]

https://collider.shinyapps.io/ColliderApp/



Notes. Panel A is a scatter plot of 1000 simulated observations where the true effect of CSR Score on Firm Performance is -0.25 (Firm Performance = -0.25*CSR + e~(0,1)). Panel B is the same dataset with only firms that have CSR disclosures.

Figure 2: Illustration of Collider Bias in CSR Score on Firm Performance

[9] 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.

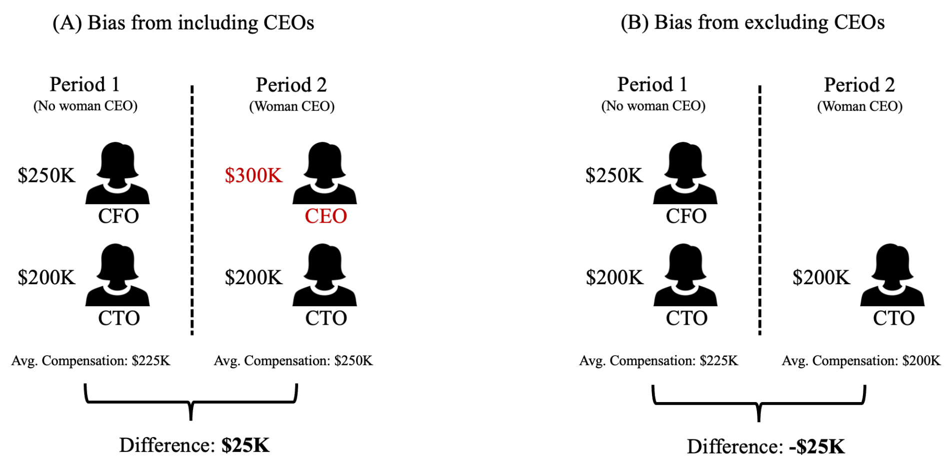


Figure 3: Illustrative example of the challenge of estimating the effect of a Woman CEO on TMT women's compensation.

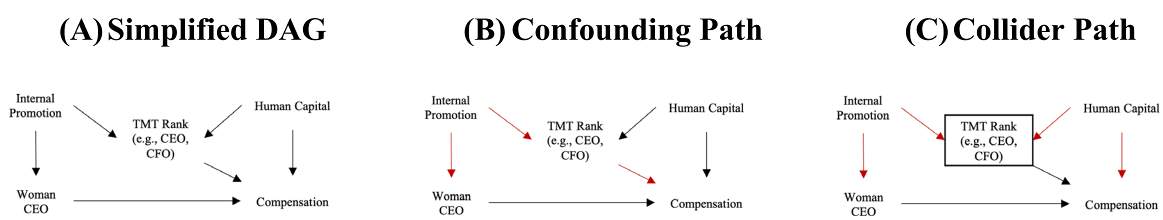


Figure 4. Simplified DAG for the effect of a woman CEO on women's compensation

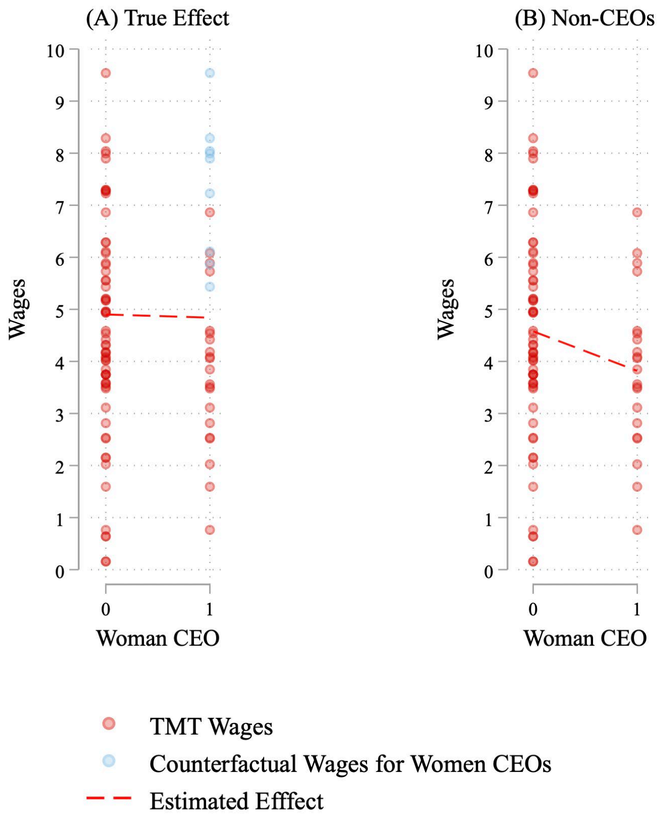


Figure 5: Illustration of Collider Bias Using Gender of CEO on the Compensation of non-CEO Women

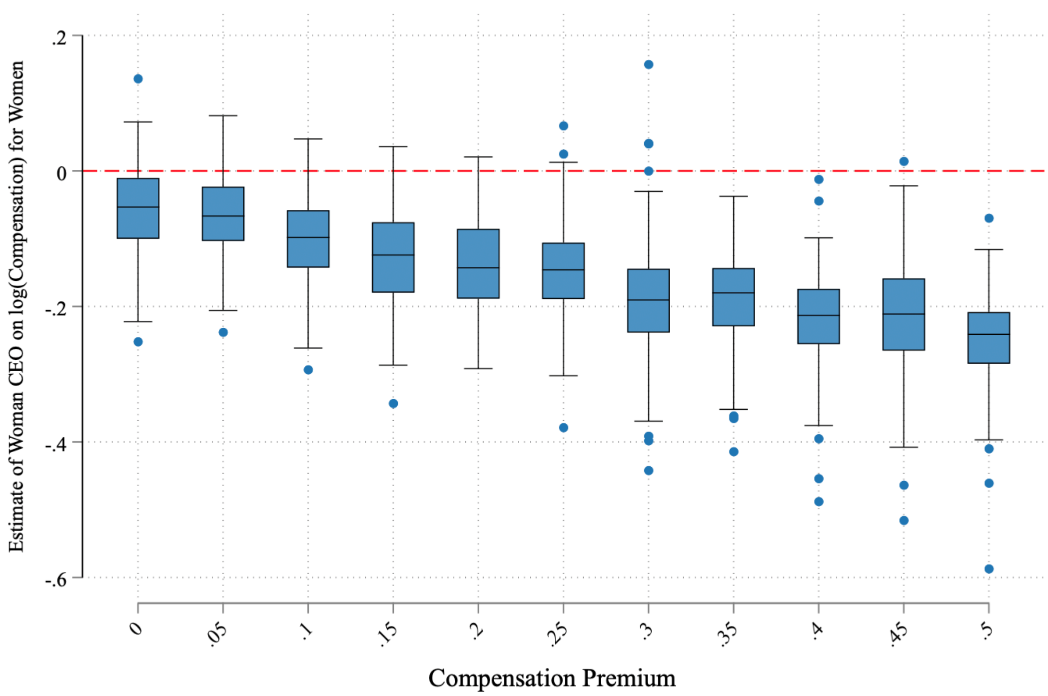


Figure 6: Estimated Effect of Women CEOs on the log(Compensation) of TMT Women, by Compensation Premium

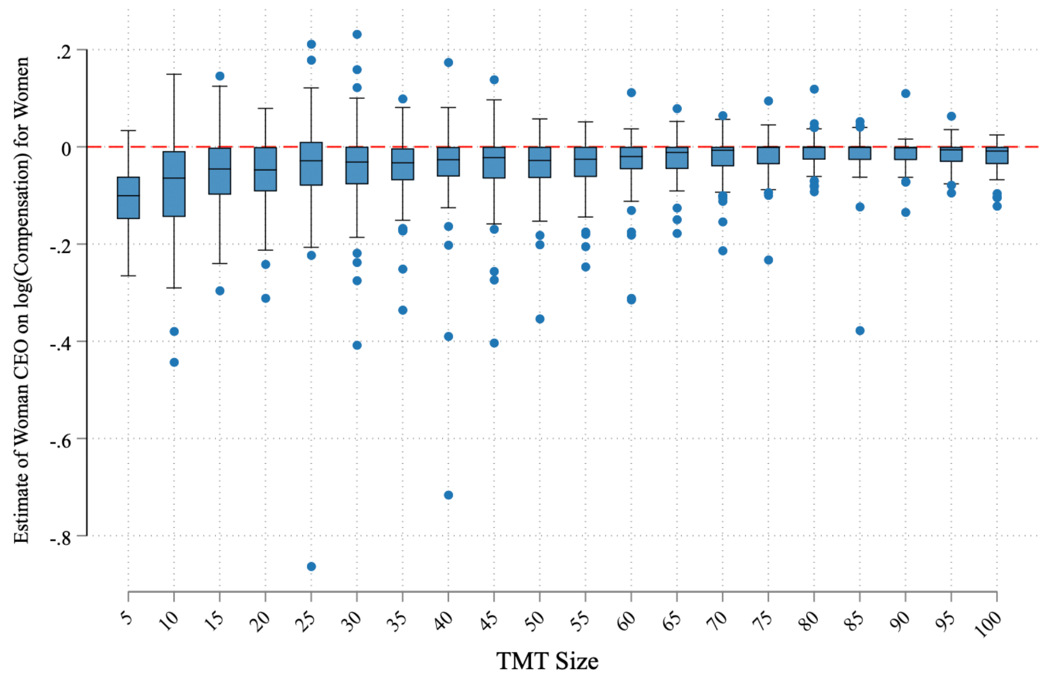


Figure 7: Estimated Effect of Women CEOs on the log(Compensation) of TMT Women, by TMT Size

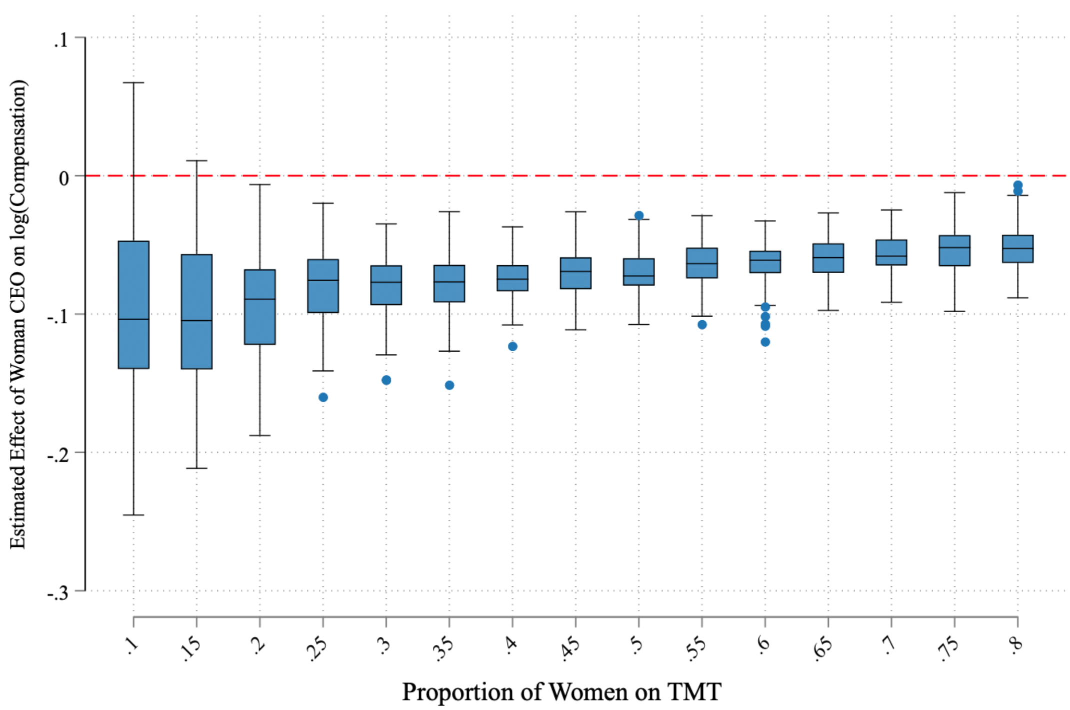


Figure 8: Estimated Effect of Women CEOs on the log(Compensation) of non-CEO Women, by Proportion of non-CEO Women on the TMT

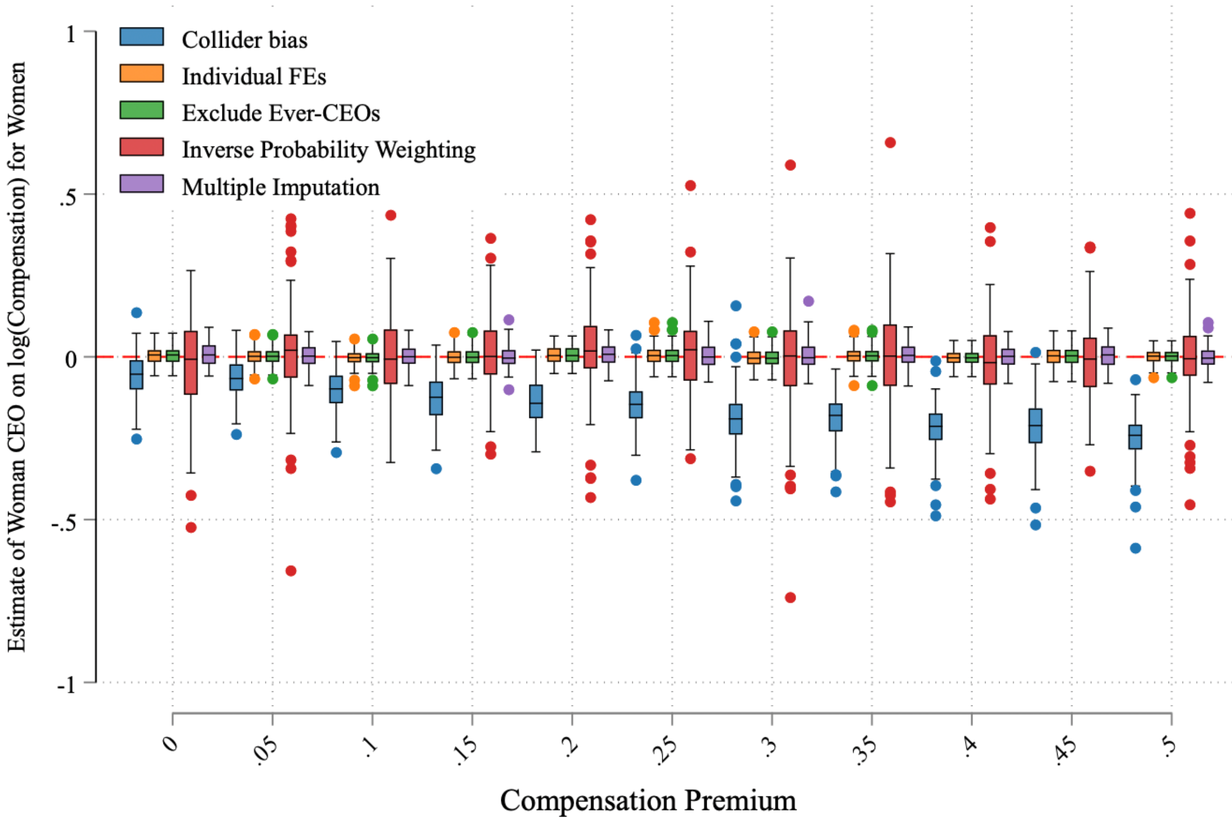


Figure 9: Estimated Effect of Women CEOs on the log(Compensation) of TMT Women, by Estimation Approach

Table 1. Frequency of Mentions of “Collider” and “Bad Controls”

Field	Journal (2010~2023)	# Published	% Discussing Colliders	% Discussing Confounders
Strategy/ Management	Strategic Management Journal	1504	1.33%	29.65%
	Management Science	1390	0.29%	13.60%
	Organization Science	741	0.54%	14.17%
	Administrative Science Quarterly	623	0.48%	13.32%
	Academy of Management Journal	617	0.32%	16.69%
	TOTAL Strategy/Management	4875	0.68%	18.99%
Economics	American Economic Review	2426	1.07%	11.87%
	Quarterly Journal of Economics	348	2.01%	19.54%
	Journal of Political Economy	328	1.22%	10.06%
	TOTAL ECONOMICS	3102	1.19%	12.54%
Sociology	American Sociological Review	690	3.62%	35.80%
	Annual Review of Sociology	197	2.54%	13.20%
	TOTAL SOCIOLOGY	887	3.38%	30.78%

Notes:
[a] Collider-related papers are any that mention “collider,” “bad control,” or "endogenous selection." Papers discussing confounders were identified by searching their text for "confounder."
[b] Because collider is a commonly used word outside of the econometric context, we read the relevant passages from each article that used this term to verify that it was used in the context of a collider variable. We also selected a random sample of articles mentioning “bad control” and “endogenous selection” and confirmed over 90% were using the terms in the context of their impact on estimation or causal inference.

Table 2. Frequency of Women and Placebo Groups in Sample

	Observations	Individuals	No. of firms
Employee is a woman	22,166	5,309	2,514
	7.4%	**9.6%**	**63.5%**
Employee first name is "John"	14,813	2,680	1,931
	5.0%	**4.9%**	**48.8%**
Employee name starts with "M"	26,893	5,003	2,712
	9.0%	**9.1%**	**68.5%**
Employee in random group	23,493	4,374	2,600
	7.9%	**7.9%**	**65.7%**

Note: Percentage of total in bold, gray background

Table 3. Summary Statistics and Correlations (N = 298,975)

		Mean	SD	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1.	Total Compensation	7.19	1.14													
2.	CEO is a woman	0.03	0.17	0.03												
3.	CEO's first name is John	0.05	0.22	0.01	-0.04											
4.	CEO's name starts with M	0.08	0.27	0.03	0.05	-0.07										
5.	CEO belongs to random group	0.07	0.26	0.02	-0.01	0.02	-0.02									
6.	Advertising Intensity	0.01	0.04	0.01	0.05	-0.02	0.04	-0.00								
7.	Firm age	3.03	0.77	0.20	0.02	0.02	0.00	0.01	-0.07							
8.	Book Leverage	0.25	0.45	0.04	-0.00	-0.00	0.00	0.01	-0.01	0.05						
9.	R&D Intensity	0.03	0.07	-0.04	-0.01	0.00	-0.00	-0.01	0.06	-0.18	-0.06					
10.	Size from assets	7.56	1.92	0.52	0.01	0.03	0.01	0.02	-0.11	0.43	0.08	-0.31				
11.	Size from employees	1.89	1.30	0.41	0.02	0.02	0.01	0.01	0.05	0.38	0.05	-0.21	0.59			
12.	Tobin’s Q	0.99	0.37	0.08	0.00	0.01	0.01	0.00	0.18	-0.15	-0.02	0.41	-0.27	-0.05		
13.	Executive's age	3.97	0.14	0.17	0.00	0.01	-0.01	0.00	-0.05	0.20	0.01	-0.07	0.16	0.11	-0.06	
14.	Employee is CFO	0.11	0.31	0.03	0.03	-0.01	0.02	0.00	-0.00	0.04	0.01	-0.01	0.04	-0.02	-0.01	-0.05

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 4. CEO Effect on the Compensation of Other Top Managers

Sample restricted to:	Other Women in TMT	Other "Johns" in TMT	Other "Starts with M" in TMT	Other "Random" in TMT
	Model 1	Model 2	Model 3	Model 4
Advertising intensity	0.097	0.496	-0.352	-0.229
	(0.806)	(0.010)	(0.398)	(0.635)
Firm age	-0.045	-0.135	-0.036	-0.153
	(0.355)	(0.058)	(0.455)	(0.001)
Leverage	-0.112	-0.223	-0.146	-0.148
	(0.036)	(0.004)	(0.001)	(0.017)
R&D intensity	-0.421	-0.498	-0.527	-0.616
	(0.066)	(0.014)	(0.090)	(0.011)
Size from assets	0.120	0.106	0.124	0.127
	(0.000)	(0.000)	(0.000)	(0.000)
Size from employees	0.133	0.171	0.144	0.140
	(0.000)	(0.000)	(0.000)	(0.000)
Tobin's Q	0.319	0.408	0.369	0.345
	(0.000)	(0.000)	(0.000)	(0.000)
Executive's age	0.425	0.654	0.417	0.334
	(0.000)	(0.000)	(0.000)	(0.001)
Employee is CFO	0.119	0.180	0.104	0.131
	(0.000)	(0.000)	(0.000)	(0.000)
CEO is a woman	-0.131			
	(0.001)			
CEO's first name is John		-0.198		
		(0.002)		
CEO's name starts with M			-0.110	
			(0.002)	
CEO belongs to random group				-0.161
				(0.000)
Constant	2.996	2.607	3.271	3.793
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	20,240	11,743	22,037	19,050
Adj. R-squared	0.707	0.707	0.657	0.670

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

Table 5. CEO Effect on TMT Representation

Dependent variable is:	% Women in TMT	% "John" in TMT	% "Starts with M" in TMT	% Random in TMT
	Model 1	Model 2	Model 3	Model 4
Advertising intensity	-0.002 (0.970)	-0.007 (0.831)	0.044 (0.473)	-0.074 (0.073)
Firm age	-0.024 (0.000)	-0.000 (0.926)	-0.011 (0.086)	-0.003 (0.586)
Leverage	0.001 (0.606)	-0.001 (0.170)	-0.001 (0.377)	0.001 (0.563)
R&D intensity	-0.005 (0.871)	-0.013 (0.544)	0.005 (0.856)	-0.045 (0.067)
Size from assets	0.000 (0.817)	-0.002 (0.167)	-0.001 (0.514)	0.001 (0.328)
Size from employees	-0.002 (0.489)	0.001 (0.721)	0.001 (0.723)	0.002 (0.540)
Tobin's Q	0.002 (0.619)	-0.001 (0.813)	0.004 (0.376)	-0.001 (0.718)
CEO is a woman	-0.031 (0.007)			
CEO's first name is John		-0.020 (0.001)		
CEO's name starts with M			-0.038 (0.000)	
CEO belongs to random group				-0.020 (0.000)
Constant	0.070 (0.000)	0.078 (0.000)	0.080 (0.000)	0.085 (0.000)
Observations	49,252	49,252	49,252	49,252
Adj. R-squared	0.462	0.393	0.406	0.403

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

Table 6. CEO Effect on Manager Compensation with and without Collider Bias

Dependent variable includes		Women	First name is "John"	Name starts with "M"	In random group
	All TMT members (p-value)	-0.131 (0.001)	-0.198 (0.002)	-0.110 (0.002)	-0.161 (0.000)
	TMT members who never become CEO (p-value)	0.012 (0.744)	0.030 (0.608)	0.012 (0.721)	0.040 (0.307)
	Percent obs. excluded within category (No. obs. excluded)	4.9% (982)	12.0% (1,414)	11.1% (2,444)	10.9% (2,068)

Notes: All models include year and firm fixed effects as well as the controls as those in table 4. P-values in parentheses. Robust standard errors clustered by firm.

Table 7. CEO Effect on TMT Representation with and without Collider Bias

Dependent variable includes:		Women	First name is "John"	Name starts with "M"	In random group
	All TMT members (p-value)	-0.031 (0.007)	-0.020 (0.001)	-0.038 (0.000)	-0.020 (0.000)
	TMT members who never become CEO (p-value)	0.012 (0.256)	0.007 (0.146)	-0.001 (0.877)	0.004 (0.416)

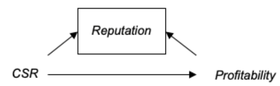
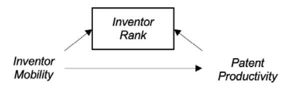
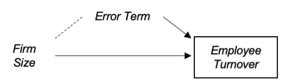
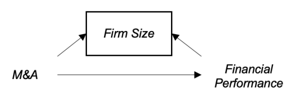
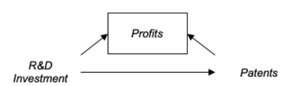
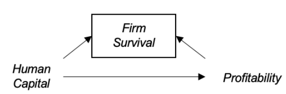
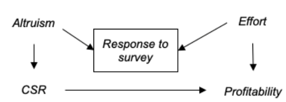
Notes: All models include year and firm fixed effects as well as the controls as those in table 5. P-values in parentheses. Robust standard errors clustered by firm.

Table 8. Monte Carlo Regressions

	(1)	(2)	(3)	(4)	(5)
DV: Log(Compensation)	Unadjusted	Only Exclude CEOs	Only Individual FEs	Exclude CEOs + Ind. Fes	Exclude Ever CEOs
Woman CEO	0.0479	-0.300	0.0469	0.00252	0.00252
	(0.000)	(0.000)	(0.000)	(0.508)	(0.814)
Bias	Positive	Negative	Positive	None	None
Sample	All Women	Non-CEO Women	All Women	Non-CEO Women	Never CEOs
Firm FE	Y	Y	Y	Y	Y
Period FE	Y	Y	Y	Y	Y
Individual FE	N	N	N	Y	N
Observations	19,856	16,529	19,856	15,812	15,812

P-values in parentheses. The true effect of Woman CEO on Log(Compensation) is 0.

Table 9. Main Forms of Collider Bias

Category	Description	Example Research Question	Simplified DAG	Potential Cause of Collider Bias	Explanation
Controls and Matching	Do not control for or match on variables that may be colliders	What is the effect of CSR on profitability?		Controlling for or matching on firm reputation	Researchers should not control for or match on a firm's concomitant reputation when studying the impact of CSR on firm profits because both CSR and profits likely affect a firm's reputation.
Fixed Effects	Do not include fixed effects that may be colliders	What is the effect of inventor mobility on patent productivity?		Inventor's rank fixed effects	Both inventor mobility and patent productivity may affect an inventor's rank in the firm (engineer vs. senior engineer), making it a collider variable when used as a fixed effect.
Selection on the Dependent Variable	Ensure that the criteria for being included in a dataset is not based on the value of the dependent variable	What is the effect of firm size on employee turnover?		Selection into making LinkedIn profile public	Employees make their LinkedIn profiles public when they would like to change employers (turnover). By using public LinkedIn profiles to measure turnover, researchers are implicitly selecting on the dependent variable.
Selection into Archival Dataset	Ensure that the criteria for being included in a dataset is not based on a collider	What is the effect of M&A on a firm's financial performance?		Selection of Fortune 500 firms	Both the M&A activity and firm performance could influence whether a firm is on the Fortune 500 list. Thus, researchers should not use a dataset of Fortune 500 firms to study the research question because being on the Fortune 500 list is a collider.
Sample Exclusion Criteria	Do not exclude observations from a sample based on a collider	What is the effect of R&D investment on patent productivity?		Excluding observations from sample based on missing measure of profit	If R&D investments and patenting both affect a firm's profitability, excluding observations where profit measures are missing would result in collider bias.
Attrition	Be aware of units that leave sample based on a collider	What is the effect of entrepreneurial human capital on startup's profitability		Exit from the sample based on firm failure	Entrepreneurs with high human capital are more likely to close their firm due to better outside options. Firms with low profits are also more likely to fail. Thus, attrition due to failure is likely a collider.
Nonresponse	Ensure that subjects do not decline responding to a survey based on a collider	Does CSR affect profitability?		Nonresponse to a survey of CEOs	If altruistic CEOS are more likely to respond to a survey and to engage in CSR, and if CEOs of high performing firms are less likely to respond to a survey because they are busy, then nonresponse is a collider.

Online Appendix A: Primer on Directed Acyclic Graphs (DAG)

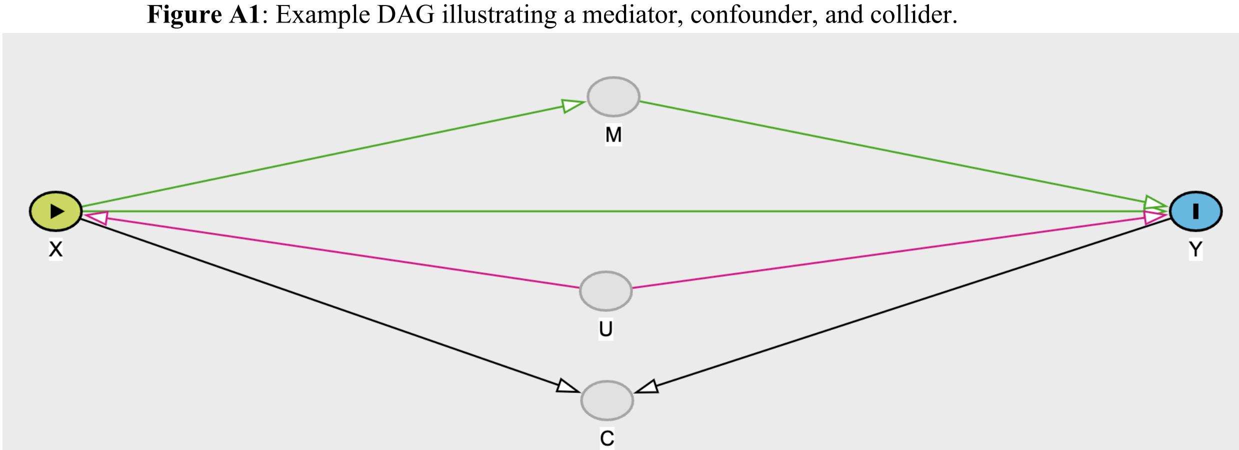
A DAG visually maps the causal relationships among variables using nodes and directed arrows. The term "acyclic" indicates the absence of cycles—paths that loop back to the starting node. Compared to other econometric approaches to causal inference like potential outcomes and structural equations, DAGs can be more accessible because they rely on graphical rules instead of mathematics. Thus, even scholars who are less fluent in mathematics can use DAGs to formally and rigorously identify potential bias in their specifications and develop strategies for mitigating bias.

DAGs allow researchers to graphically explicate their theoretical model. In a DAG, nodes represent random variables, which can either be observed or unobserved. Arrows in the DAG represent direct causal relationships. Just as the presence of an arrow represents a causal relationship, the absence of arrows between variables signifies the assumption that no causal relationship exists between them.

A "path" refers to an ordered series of arrows linking two variables, without regard for the arrow direction. Each variable can only be traversed once on a given path.

When all arrows on a path point from the treatment variable to the outcome variable, it is referred to as a causal path. In general, causal identification aims to identify total effects across all causal paths. For example, in Figure A1 the total causal effect of the treatment (X) on the outcome (Y) would include the following causal paths: X → Y and X → M → Y. Any other paths that are not causal, for example X ← U → Y and X → C ← Y are referred to as "biasing paths."

Figure A1: Example DAG illustrating a mediator, confounder, and collider.



There are three basic causal structures that make up any DAG. The first is called a chain, where all arrows are oriented in the same direction. Chains represent causal paths. In Figure A1, X→Y and X→M→Y are causal paths. A variable that falls between the explanatory and outcome variables in a causal chain is referred to as a mediator. Conditioning on a mediator (M) blocks the causal path, which induces overcontrol bias.

The second basic structure is called a fork, where one variable directly causes two or more other variables. In Figure A1, X←U→Y is a fork. U is referred to as a confounder. Conditioning on U eliminates this spurious association, called confounding bias or omitted variable bias, that would otherwise exist between X and Y.

The third structure is referred to as an inverted fork, which is depicted in Figure A1 by X→C←Y. Here, X and Y both cause C, known as a collider. As opposed to a confounder, conditioning on a collider (C) creates a spurious association between X and Y, called "collider bias." In summary, to identify causal effects researchers should condition only on confounders, not on mediators or colliders.

Essentials for Creating a DAG

To construct a DAG, you will need:

- A comprehensive understanding of the research variables.
- Expertise or relevant literature that outlines the variable relationships.
- A tool for sketching the DAG, like specialized software or basic pen and paper.

Steps to Construct a DAG

To construct a DAG for a specific research question, researchers can follow the following steps:

Step 1: Clarify Research Question and Variables

Begin by detailing your research question, defining the unit of analysis and identifying dependent and independent variables. For instance, in formulating the research question from the empirical example in Section 2.3 of the main paper, the question is: "what is the effect of CEO gender on the pay of women top-management team (TMT) members?" Here, the unit of analysis is the TMT member-year.

Step 2: Recognize Immediate Causal Factors

List the variables directly affecting your independent and dependent variables. Use your domain knowledge and existing studies for this. For example, if your question focuses on the effect of CEO gender on women TMT members' compensation, relevant variables might include individual-level factors like human capital, rank, age, effort, and tenure, and firm-level elements like profitability, stock price, and firm size.

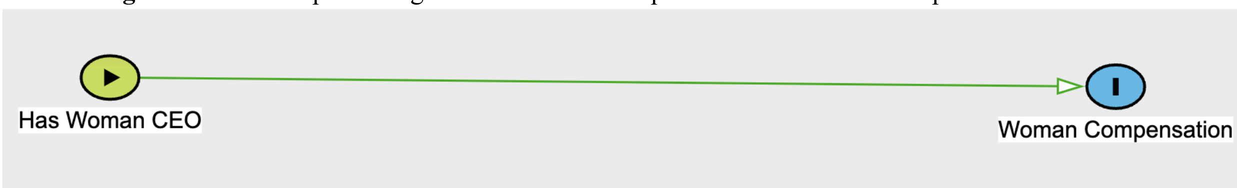
Step 3: Sketch the DAG

With your variables identified, start drawing the DAG as follows:

a) **Choose a DAG drawing tool**: Many software tools are available for drawing DAGs, such as DAGitty, Lucidchart, or even a pen and paper. A complete tutorial on how to use the various software for drawing DAGs is beyond the scope of this primer. For this exercise, we used an online tool at daggity.net, which we recommend as an easy and powerful tool for creating DAGs. For help using daggity.net, we point the interested reader to the manual at https://www.dagitty.net/manual-3.x.pdf.

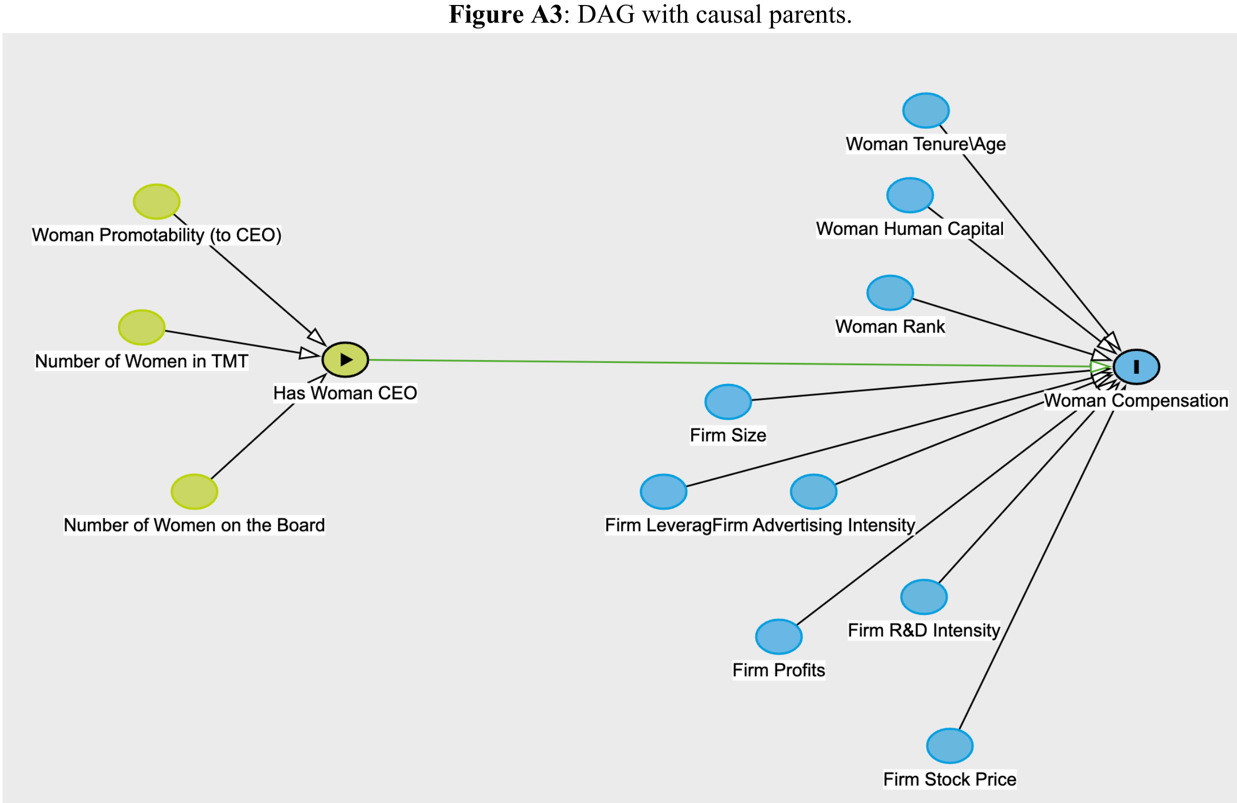
b) **Map the Independent and Dependent Variables**: Draw your independent and dependent variables and connect them with an arrow indicating causality. Various tools have different symbols for representing these variables, but the general idea remains the same: nodes represent variable and arrows signify causal relationships. As seen in Figure A2, daggity.net represents the independent variable, *Has Woman CEO*, with a green oval around a triangle and the dependent variable, *Woman Compensation*, with a blue oval around a bar. The causal path of interest is represented with a green arrow.

Figure A2: DAG representing the effect of the independent variable on the dependent variable.



c) **Add Causal Parents:** After mapping the independent and dependent variables, draw and connect all causal parents identified in Step 2. In Figure A3, the causal parents of the independent variable are depicted with green ovals, while those of the dependent variable are in blue ovals.

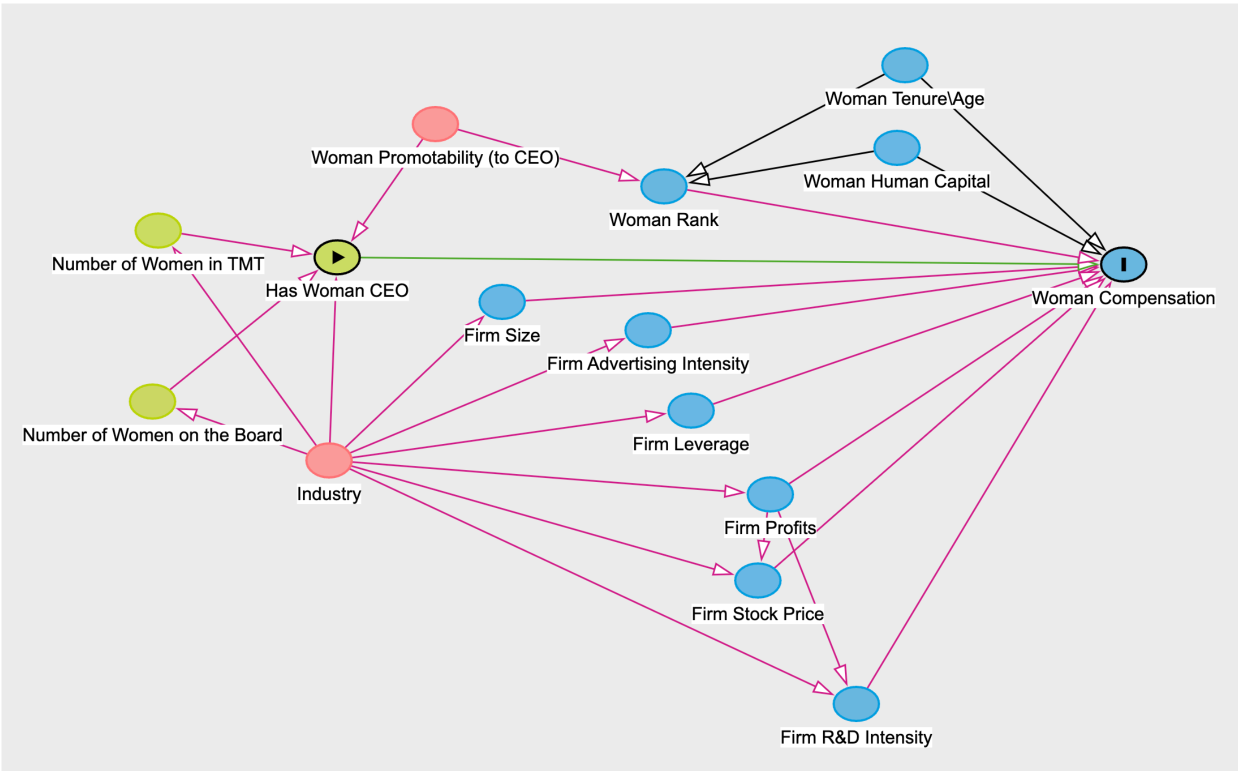
Figure A3: DAG with causal parents.



d) **Expand the Causal Map**: Now, think about other variables that could either cause or be caused by the already-listed variables. Add these to the graph and connect them with arrows to signify their relationships. Figure A4 represents the authors' initial attempt to visually represent the causal map of the research question. For example, we added the variable

Woman Promotability, which refers to the unobserved opportunity structure at a firm by which a focal woman may come to occupy the CEO position, such as the previous CEOs retirement or firing, and the availability of other suitable candidates.

Figure A4: DAG without controls or selection.



Daggity highlights all biasing paths—often termed "backdoor paths" or "bad paths"—by coloring the arrows pink.

Step 4: Refine the DAG: It is unlikely that the first draft of the DAG will be a satisfactory representation of the author's causal map. To refine the DAG:

a) Validate Causal Links: Confirm that all arrows denote causality, not just correlations. If you find non-causal arrows, it usually indicates a missing variable or set that should be added to clarify the causal structure.

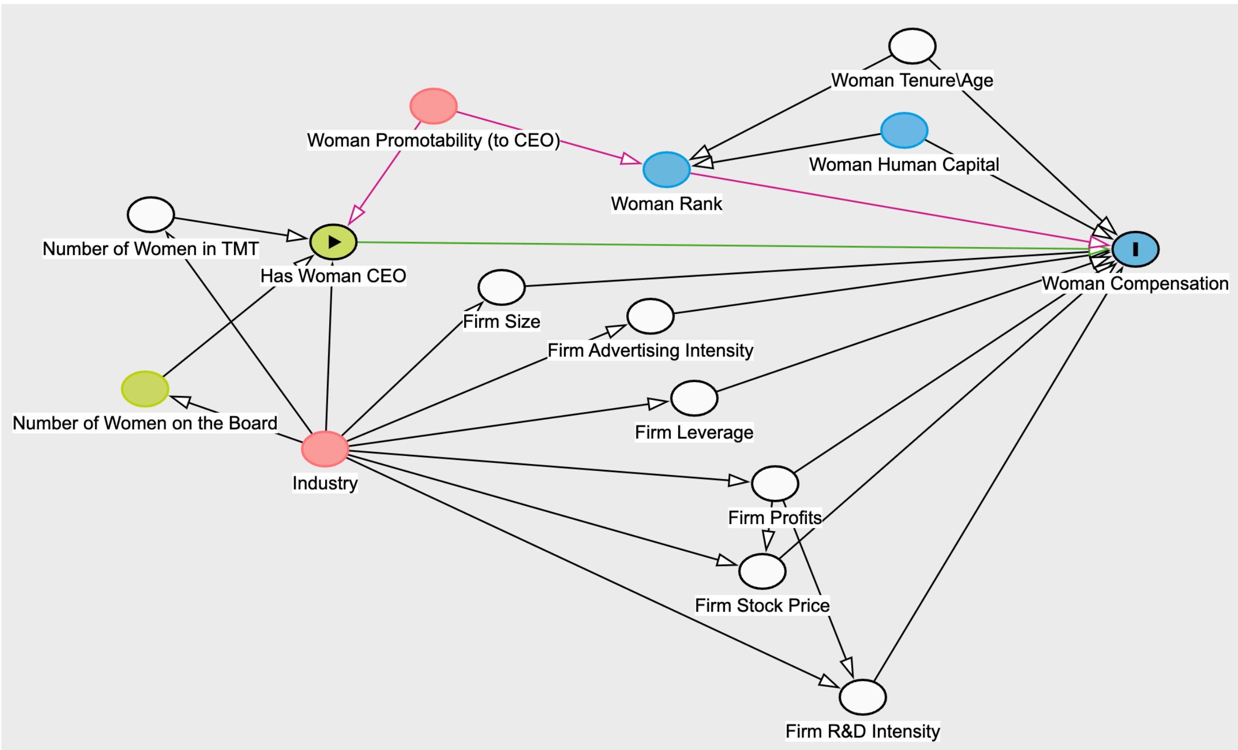
b) Eliminate Cycles: Remember, a DAG should not have any loops. Review the graph to ensure it remains acyclic.

c) Peer Review: Share your initial DAG with peers or domain experts to confirm the validity of the causal relationships. This feedback can help you revise and improve your DAG.

Step 5: Close Biasing Paths

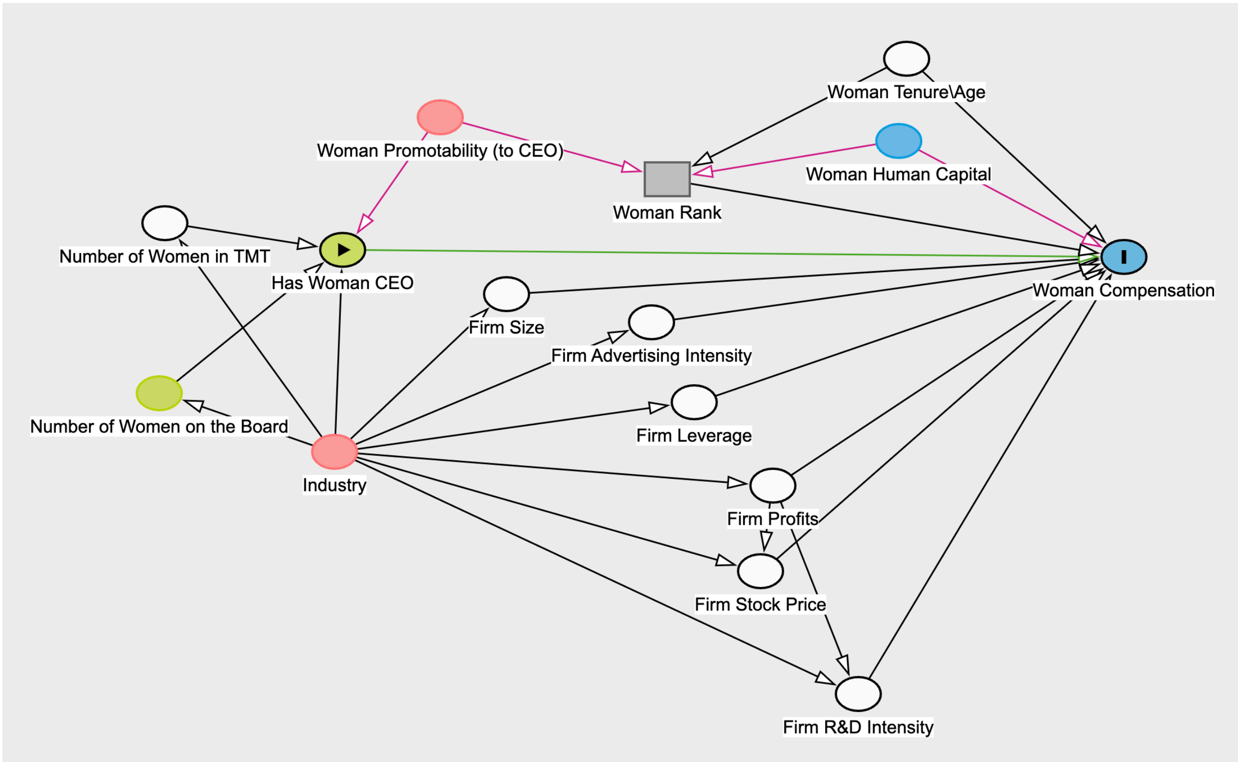
When defining variables in Daggity, the tool assumes the variables are not conditioned on (controlled or selected on) by default. You can mark the controlled and selected variables, which Daggity represents with white ovals and grey rectangles, respectively. Importantly, conditioning on confounders closes the biasing path, while conditioning on colliders opens biasing paths. For example, after indicating the controls used by DLR, Daggity would produce the DAG in Figure A5.

Figure A5: DAG with controls from DLR



Take note that these controls successfully closed many biasing paths (formerly pink arrows are now black) by blocking confounders. However, we still observe a biasing path—represented by pink arrows—from *Has Woman CEO* ← *Woman Promotability (to CEO)* → *Woman Rank* → *Woman Compensation*. This is because if a woman is highly promotable (to CEO) and she is promoted to CEO, it directly affects whether the firm has a woman CEO, but it also affects the prompted woman's compensation because she is likely to receive higher compensation after becoming a CEO (changed her rank). In other words, *Woman Promotability (to CEO)* is a confounder on the path *Has Woman CEO* ← *Woman Promotability* → *Woman Rank* → *Woman Compensation*. For this reason, DLR condition on Woman Rank by excluding (selecting) TMT women in the CEO position. We can indicate this in Daggity, which represents selection on a variable by drawing a grey rectangle. The resulting DAG is seen in Figure A5.

Figure A6: DAG showing how selecting on *Woman Rank* opens a biasing path.



Notice, this successfully closed the confounding path seen in Figure A5 from *Woman Promotability* → *Woman Rank* → *Woman Compensation*. Specifically, the path from *Woman Rank* → *Woman Compensation* is no longer pink, meaning it is no longer a biasing path.

However, *Woman Rank* is also a collider according to the DAG. Specifically, it is a collider on the following path: *Has Woman CEO* ← *Woman Promotability* → *Woman Rank* ← *Woman Human Capital* → *Woman Compensation*. Thus, while conditioning on *Woman Rank* closed one biasing path by conditioning on a confounder, it opened another biasing path by conditioning on a collider. Figure A5 graphically represents the main models presented by DLR and helps illustrate the fundamental problem of collider bias in that paper.

To effectively close the remaining bias path in Figure A6, we can either condition on *Woman Promotability (to CEO)* or *Woman Human Capital*. One specific approach to conditioning on *Woman Promotability* is to exclude TMT women who are ever promoted to CEO from the sample. Relatedly, researchers could condition on the woman's human capital by including individual fixed effects (assuming within-person human capital is stable over the sample period). As shown in Figures A7 and A8, both approaches effectively close the biasing path caused by conditioning on the collider *Woman Rank*.

Figure A7: DAG showing how conditioning on *Woman Promotability (to CEO)* closes the biasing path in Figure A6.

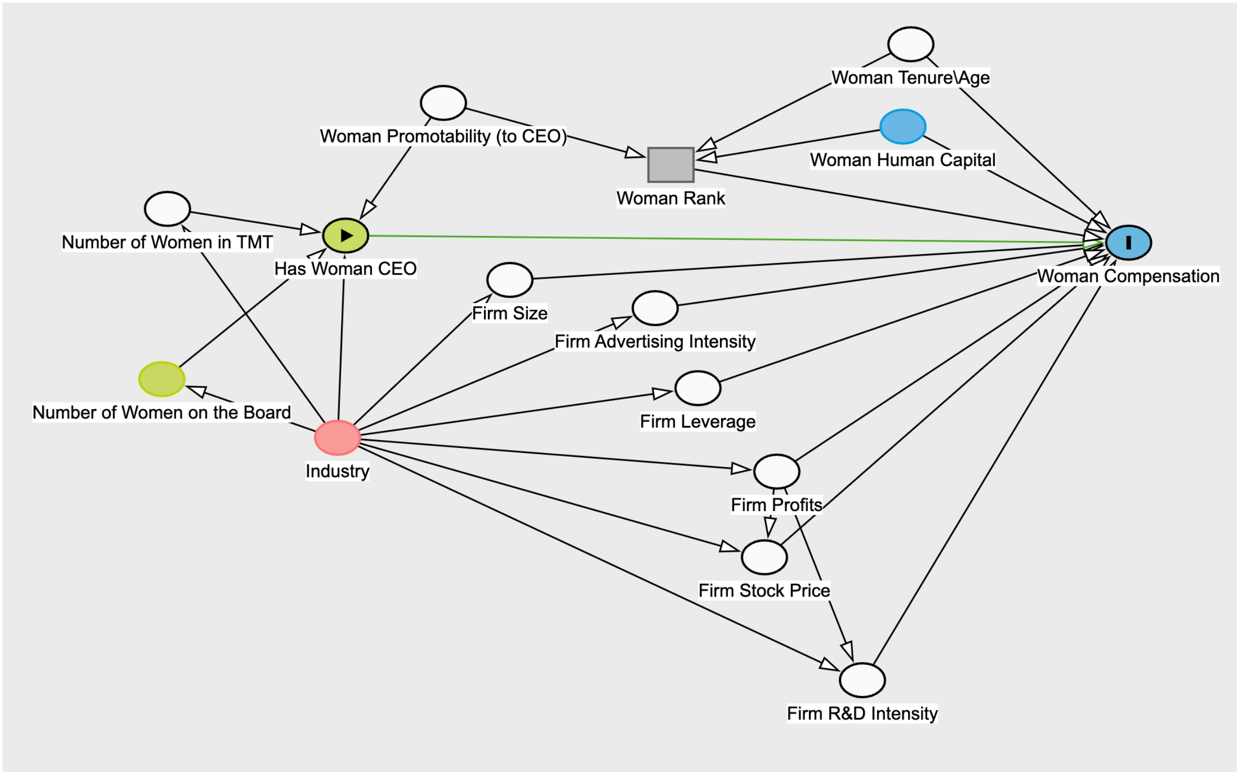
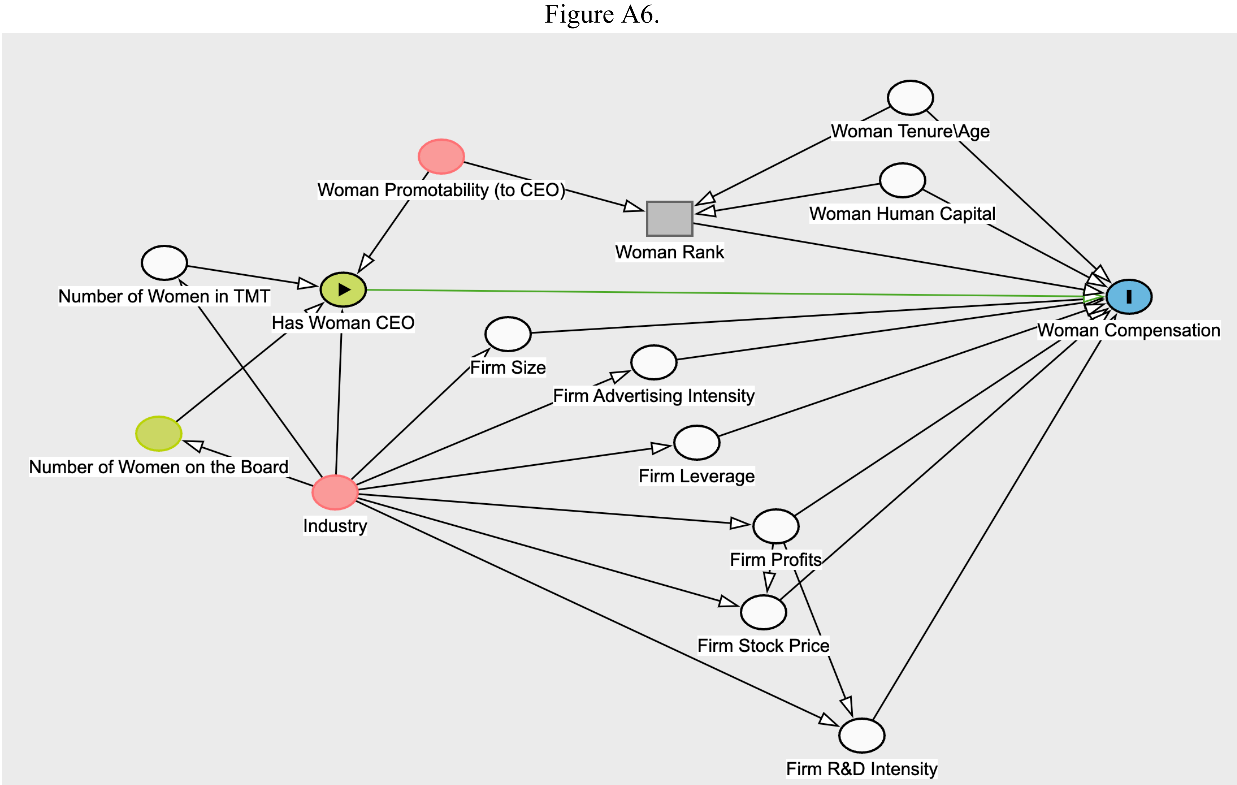


Figure A8: DAG showing how conditioning on *Woman Human Capital* closes the biasing path in Figure A6.



In conclusion, Directed Acyclic Graphs (DAGs) are a powerful tool for illustrating the researchers assumed causal map of the research question, for identifying potentially biasing paths, and for understanding how conditioning on variables is likely to affect causal inference. Thus, we encourage researchers to use DAGs to identify a remedy threats to causal inference, such as confounding and collider bias.

Online Appendix B: Replication of DLR and CLR main effect

Conventional replications are of limited use in detecting and addressing collider bias that arises from endogenous selection. In the current application, collider bias yields misleading results, but these results persist across different sampling schemes and are robust to the inclusion (or exclusion) of time-varying control variables. We demonstrate this point by conducting a replication of the principal results in DLR and in CLR. To assess the robustness of their main result, we varied four common forking paths according to King et al. (2020); sample selection in terms of executives and firms included in the analysis, years included in the sample, and the inclusion/omission of a set of time-varying control variables. In both cases the various coefficients estimated support the incorrect conclusion that having a woman CEO has negative effects on the career outcomes of other executive women.

A.1 Replicating the DLR Effect

Our estimation strategy is the same as that deployed by DLR, who use OLS regressions with firm and year fixed effects. We followed variable construction and sampling closely to the descriptions found in DLR. Thus, our exercise can be categorized as a narrow replication (Bettis et al, 2016). Table A1 below displays the coefficients, standard errors and choices for 16 models in comparison to the result reported in the DLR article (in Table 5, Model 2, under the "Women" column). The models are sorted in ascending order by magnitude of the effect.

Table A1: Models Replicating the DLR Effect of *Has Woman CEO* on *Other Women's Compensation*

			Executives selected		Firms selected		Years in sample		Additional controls	
MODEL	Coeff.	SE	CEO + top 4	All reported	Only S&P 1500	All firms in ExecuComp	1992-2013	1992-2021	No controls	Time-varying controls
A	-20.49	5.1	x			x		x	x	
B	-19.61	5.91	x			x	x		x	
C	-19.06	5.78	x			x	x			x
D	-18.19	3.97	x			x		x	x	
E	-18.09	5.35		x	x			x	x	
F	-17.61	9.75	x		x		x		x	
G	-17.47	9.89	x		x		x			x
H	-17.14	4.83	x			x		x		x
DLR	-16.24	5.74	x			x	x			x
I	-16.08	3.91	x			x		x		x
J	-15.95	5.85		x		x	x		x	
K	-15.91	5.61		x		x	x			x
L	-15.13	4.84		x	x			x		x
M	-14.94	4.09		x		x		x	x	
N	-14.45	9.5		x	x		x			x
O	-14.02	9.5		x	x		x		x	
P	-13.12	3.83		x		x		x		x

Notes:

[a] All models have separate firm and year fixed effects, and standard errors clustered by firm

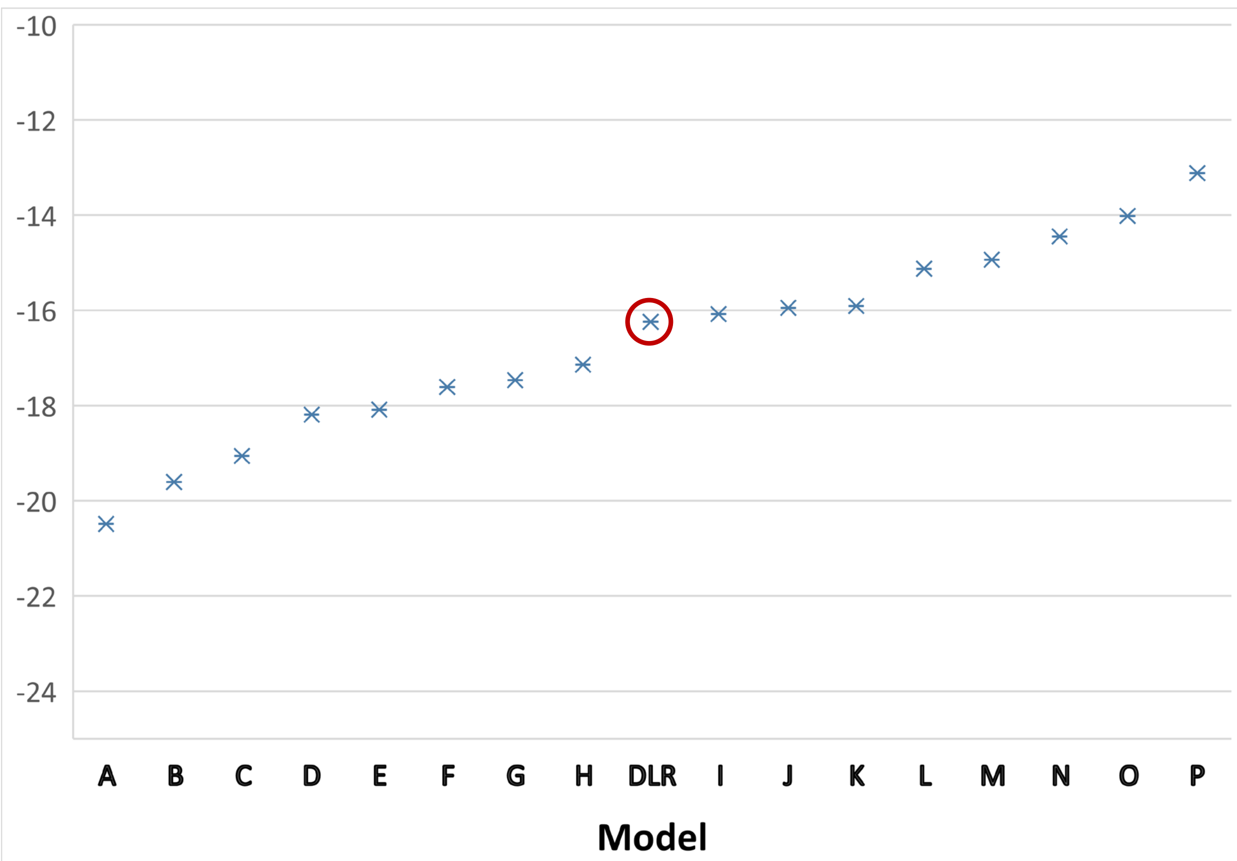
[b] Seven time-varying controls firm-level controls and two executive-level controls are described in section 3.3.3 of the main paper

[c] Coefficients have been multiplied by 100 to conform with DLR

Based on our analysis of DLR, model C in table A1 is the closest to their model in terms of selection of executives and firms, years in the sample, and most control variables.[10]

All models yielded large negative effects, between -13% and -20% compensation penalty for having a woman CEO compared to a man CEO. Most coefficients for *Has Woman CEO* were precisely estimated with the exception of noisy coefficients for models N and O. As can be seen in figure A9 below, the effect found DLR is well within the range of coefficients estimated. In the main analyses in our paper, we use model P (the most conservative model), which extends the sample to include a greater number of executives that report compensation data and adds eight more years to encompass the 1992-2021 time period.

Figure A9: Range of Coefficient of *Has Woman CEO* on *Other Women's Compensation*



Notes:
Coefficients estimated using the models described in Table A1

[10] DLR's models include a longer list of controls for manager's job titles, which requires an additional dataset (BoardEx), but which should have no bearing on collider bias because these would not block the problematic pathways in the DAG.

A.2 Replicating the CLR Effect

In order to be consistent with the DLR application, we used OLS regressions with firm and year fixed effects. This estimation strategy differs from that employed by CLR, who use a random effects Tobit estimator. In addition, we do not include various CEO characteristics in our models, and a different set of firm-level covariates. Third, CLR 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). Given differences in both modelling and sampling, our exercise is akin to a pseudo replication (Bettis et al, 2016).

We compare results from our 16 models to the main effect found by CLR and reported in their paper in Table 2 (model 2). Model E in Table A2 is the closest to the CLR model in terms of selection of executives and firms, years in the sample, and the use of time-varying control variables.

All models yielded stable negative effects, ranging between -3% and -7% TMT gender diversity penalty for having a woman CEO compared to a man CEO. All coefficients for *Has Woman CEO* were precisely estimated. As can be seen in figure A10 below, the effect found in CLR is within the range of coefficients estimated. In the main analyses in the paper, we use model O, which extends the sample to include a greater number of executives that report compensation data and adds eight more years to encompass the 1992-2021 time period.

Table A2: Models Replicating the CLR Effect of *Has Woman CEO* on *% Women in TMT*

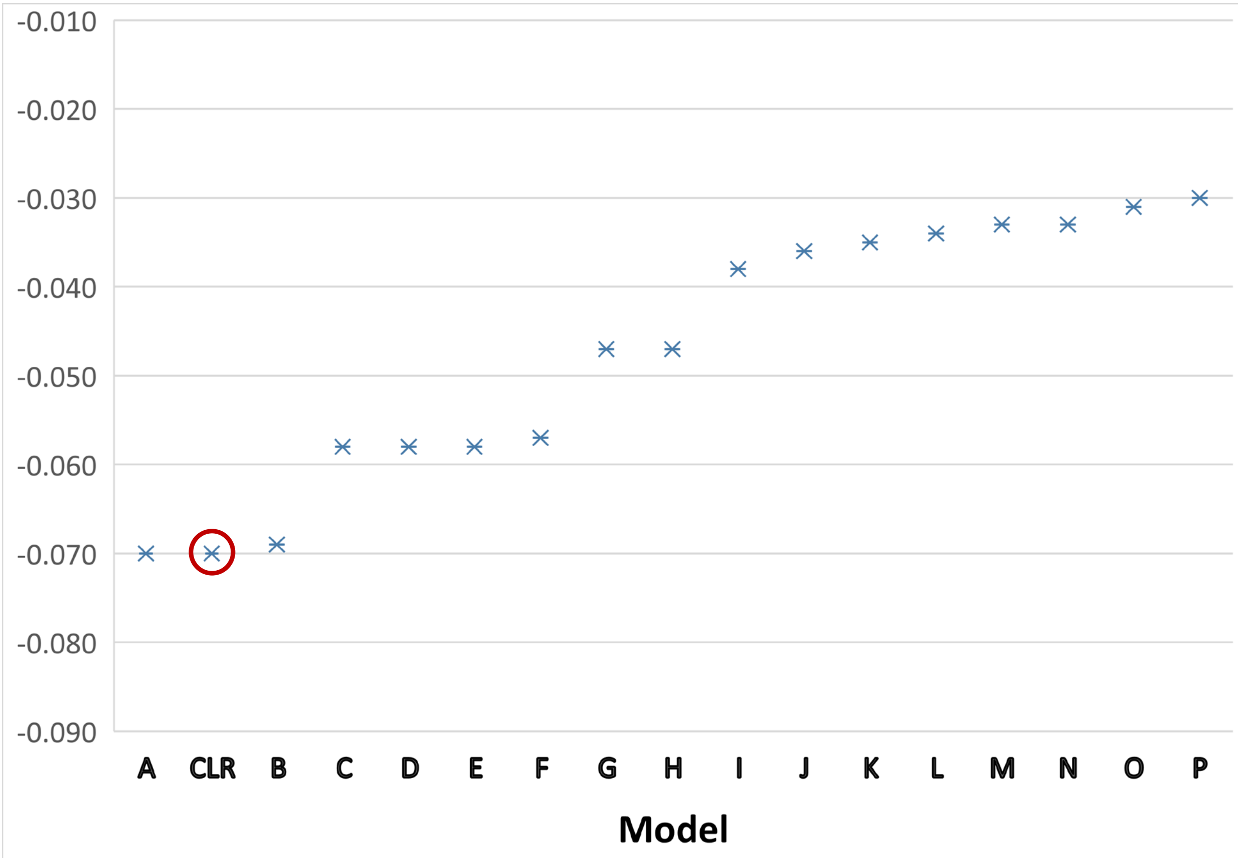
			Executives selected		Firms selected		Years in sample		Additional controls	
MODEL	Coeff.	SE	CEO + top 4	All reported	Only S&P 1500	All firms in ExecuComp	2011-2017	1992-2021	No controls	Time-varying controls
A	-0.070	0.022	x		x		x			x
CLR	-0.070	0.030		x	x		x			x
B	-0.069	0.021	x		x		x		x	
C	-0.058	0.018	x			x	x			x
D	-0.058	0.018	x			x	x		x	
E	-0.058	0.021		x	x		x			x
F	-0.057	0.021		x	x		x		x	
G	-0.047	0.017		x		x	x			x
H	-0.047	0.017		x		x	x		x	
I	-0.038	0.016	x		x			x		x
J	-0.036	0.016	x		x			x	x	
K	-0.035	0.016		x	x			x		x
L	-0.034	0.012	x			x		x		x
M	-0.033	0.012	x			x		x	x	
N	-0.033	0.016		x	x			x	x	
O	-0.031	0.012		x		x		x		x
P	-0.030	0.012		x		x		x	x	

Notes:

[a] All models have separate firm and year fixed effects, and standard errors clustered by firm

[b] Time-varying controls include seven firm-level controls described in section 3.3.3 of the main paper

Figure A10: Range of Coefficient of *Has Woman CEO* on *% Women in TMT*



Notes:
Coefficients estimated using the models described in Table A2

Online Appendix C: Individual fixed effects in the DLR application

In section 3.5 of the article, we claim that estimation using individual fixed effects is similar to dropping future CEO observations in the DLR application. Depending on whether the model has two periods or more than two periods, adding individual fixed effects will ensure that future CEO observations have either exactly zero influence or very little influence on the estimates.

1) **Two period model (zero influence on estimates)**

Suppose there are two time periods $t = 1{,}2$ and individuals are indexed by i. Consider a regression of $Y_{i,t}$ on a generic vector of controls $X_{i,t}$ including a constant,

$$Y_{i,t} = \beta^T X_{i,t} + U_{i,t} + V_i,$$

where β is an unknown parameter, the T superscript denotes transpose, $U_{i,t}$ has mean zero conditional on $(X_{i,1}, X_{i,2})$, and V_i contains individual unobserved heterogeneity that can be correlated with $X_{i,t}$. Suppose for now that we are including fixed effects for every individual, even if they are not a future CEO. By the Frisch-Waugh-Lovell theorem, individual fixed effects regression is the same as running least squares after subtracting time averages

$$Y_{i,t} - \bar{Y}_i = \beta^T (X_{i,t} - \bar{X}_i) + (U_{i,t} - \bar{U}_i),$$

where $\bar{Y}_i$ is the average of $Y_{i,t}$ for all time periods available for individual i. For instance, if both time periods are available then $\bar{Y}_i = (Y_{i,1} + Y_{i,2})/2$ and if only the first time period is available then $\bar{Y}_i = Y_{i,1}$. The $\bar{X}_i$ and $\bar{U}_i$ are defined similarly and V_i drops out of the regression because it is constant over time. Let $Z_{i,t} = X_{i,t} - \bar{X}_i$. If only one period is available for an individual, then $Z_{i,t} = 0$ for that individual. In particular, the future CEO is only available in period 1 but not in period 2 and thus has $Z_{i,t} = 0$. By the Frisch-Waugh-Lovell theorem, the fixed-effects estimate of β can now be represented as

$$\hat{\beta} = (\sum_i \sum_t Z_{i,t} Z_{i,t}^T)^{-1} \sum_i \sum_t Z\,(Y_{i,t} - \bar{Y}_i).$$

As can be seen, any $Z_{i,t}$ with value zero completely drops out of this expression, which has the same effect as not including this observation in the first place.

Now suppose individual-level fixed effects are included *only* for future CEOs. In the expression for $\hat{\beta}$, the $Z_{i,t}$ and $(Y_{i,t} - \bar{Y}_i)$ are then replaced by $X_{i,t}$ and $Y_{i,t}$ for non-CEO observations. For future CEO observations, $Z_{i,t}$ and $(Y_{i,t} - \bar{Y}_i)$ remain unchanged and $Z_{i,t}$ is still zero, which is the same as not including future CEOs not in the regression.

2) **Multiple period model (very low influence on estimates)**

Now suppose there are more than two time periods. In that case, a future CEO (or any other observation where some time periods are not available) is no longer completely dropped from the regression by a fixed effect. However, the fixed effect still heavily dampens the impact of the future CEO on $\hat{\beta}$. The reason is that the future CEO fixed effect changes $X_{i,t}$ to $X_{i,t} - \bar{X}_i$ and $Y_{i,t}$

to $Y_{i,t} - \bar{Y}_i$, which ensures that only the variation about the average level of $X_{i,t}$ and $Y_{i,t}$ enters the regression. Whether $X_{i,t}$ and $Y_{i,t}$ are particularly high or low for some observations plays no role when fixed effects are included. In the present case, the potentially very high average compensation of the future CEO prior to promotion now does not enter the regression. Only the changes about that average compensation matter. The fact that the future CEO has a high compensation causes the bias in the estimates of the effect of a female CEO that we are concerned about and an individual fixed effect effectively removes this bias.

In sum, future CEO fixed effects and dropping future CEOs is identical in a two-period regression model. Including individual fixed effects accounts for individual heterogeneity and also drops CEOs from the sample in a two-period regression model. In a longer panel, the effect of a future CEO is severely dampened by a fixed effect. In all cases, future CEOs have little to no influence on the computation of $\hat{\beta}$ when individual fixed effects are used.

The following empirical exercise helps corroborate the above explanation. We estimated 16 models replicating the DLR effect of *Has Woman CEO* on the compensation of other women based on the "forking paths" described earlier in online appendix B. In addition to firm and year fixed effects, we added individual fixed effects for every executive in all the models. Below in table A3 are the coefficients and p-values for those 16 models, along with the sampling characteristics for each model. In line with our mathematical explanation, the effect of *Has Woman CEO* on the compensation of non-CEO women goes away across all models once individual fixed effects are included.

Table A3 Coefficients of *Has Woman CEO* for models including individual fixed effects

			Executives selected		Firms selected		Years in sample		Additional controls	
MODEL	Coeff.	p-value	CEO + top 4	All reported	Only S&P 1500	All firms in ExecuComp	1992-2013	1992-2021	No controls	Time-varying controls
A	0.010	0.850		x		x		x	x	
B	0.000	0.964		x		x		x		x
C	0.030	0.554	x			x		x	x	
D	0.020	0.727	x			x		x		x
E	0.000	0.953	x		x			x	x	
F	-0.010	0.918	x		x			x		x
G	0.030	0.670	x			x	x		x	
H	0.010	0.929	x			x	x			x
I	0.010	0.932	x		x		x		x	
J	-0.010	0.903	x		x		x			x
K	-0.010	0.825		x	x			x	x	
L	-0.020	0.738		x	x			x		x
M	0.010	0.902		x	x		x		x	
N	-0.030	0.754		x	x		x			x
O	0.010	0.877		x		x	x		x	
P	-0.010	0.819		x		x	x			x

References

Bettis, R. A., Helfat, C. E., & Shaver, J. M. (2016). The necessity, logic, and forms of replication. Strategic Management Journal, 37(11), 2193-2203.

King, A., Goldfarb, B., & Simcoe, T. (2021). Learning from testimony on quantitative research in management. Academy of Management Review, 46(3), 465-488.

Online Appendix D: Annotated Stata code for DLR application

```
/*******************************************************************
Data downloaded from WRDS in December 2022
1) "compustat_controls" is the full Compustat annual file including all available firms and years,
used to create firm-level controls as explained in the paper
2) "Execucomp_main" is the full Execucomp annual file including all available firms, variables and years
**********************************************************/
clear all
set more off
use compustat_controls,clear
merge 1:m year gvkey using Execucomp_main ,clear
keep if _m==3
keep if inrange(year,1992,2021)                 //years in full sample for analysis
drop if missing(tdc1)                           //drop if no compensation info for executive
isid execid gvkey years                         //confirm data structure
replace tdc1=0 if tdc1<0                        //10 corrections
replace tdc1=ln(1+tdc1)
sort gvkey year tdc1
gen female=(gender=="FEMALE"|nameprefix=="Ms.")
replace female=0 if nameprefix=="Mr."                         //fix 65 errors
gen xx=2022-year+1
replace age =page -xx if missing(age)
gen lnage=ln(1+age)                                           //calculate ln of exec. age
qui su lnage
replace lnage=r(mean) if missing(lnage)         //Impute to conserve observations -does not affect results-
gen CEO=!missing(ceoann)                        //a dummy for CEOs
gen cfo=!missing(cfoann)                        //a dummy for CFOs
//generate 3 artificial groups for Placebo regressions
gen John=exec_fname=="John"                     //employees whose first name is "John"
gen fletM=substr(exec_fname,1,1)=="M"           //anyone whose first name starts with "M"
set seed 123
egen select = tag(execid)
gen rand = runiform() if select==1
gsort execid -select
replace rand = rand[_n-1] if execid==execid[_n-1] & select==0
gen Rgroup=(rand<=0.08)                         //a random group of 8% of employees
keep female year execid gvkey tdc1 lnage CEO Rgroup fletM John cfo
bysort gvkey execid: egen eceo=max(CEO)         //employee has ever been CEO at that firm
foreach v in female John fletM Rgroup{
gen `v'ceo=CEO & `v'                            //Employee with attribute "v" is CEO
bysort gvkey year: egen `v'CEO=max(`v'ceo)      //Firm year has CEO with attribute "v"
}
```

```
compress
label var tdc1 "Total Compensation"
label var lnage "Executive's age"
label var fletMCEO "CEO's name starts with M"
label var RgroupCEO "CEO belongs to random group"
label var JohnCEO "CEO's first name is John"
label var femaleCEO "CEO is a woman"
label var cfo "Employee is CFO"
label var tdc1 "Total Compensation"
label var booklev "Book Leverage"
label var lntq "Tobin's Q"
label var lnRDint "R&D Intensity"
label var firmage "Firm age"
label var lnAdvint "Advertising Intensity"
label var lnemp "Size from employees"
label var lnat "Size from assets"

/***************************************************************
Main Analysis of DLR results below
*******************************************************/

global Xcont lnAdvint firmage booklev lnRDint lnat lnemp lntq  lnage cfo //control variables

//Biased effect: see Table 4 in the main paper
est clear
foreach v in female John fletM Rgroup{
qui reghdfe tdc1 $Xcont i.year `v'CEO if !CEO & `v' , absorb(gvkey) cluster(gvkey)
est store M_`v'
}
//Remove attriters to unbias estimate: see Table 6 in the main paper
foreach v in female John fletM Rgroup{
qui reghdfe tdc1 $Xcont i.year `v'CEO if !eceo & `v' , absorb(gvkey) cluster(gvkey)
est store c_not`v'
}
//Individual fixed effects also unbias estimate: see online Appendix C
qui reghdfe tdc1 i.year $Xcont femaleCEO if female, absorb(gvkey execid) cluster(gvkey)
est store c_FE_id

estimates table c_*, b(%9.2f) p title(Effect of CEO attribute on non-CEO compensation) varwidth(25)
drop(i.year) varlabel stats(r2_a N N_clust)
estimates clear
```