# Who gets the benefit of the doubt?

# CEO gender and news about firm performance*

Marcela Carvalho†

November 28, 2024

I show that financial markets react asymmetrically to bad news about firm performance depending on whether it is delivered by male or female CEOs. This asymmetry manifests itself in analysts' forecasts, stock returns on earnings announcement days, and even in the tone that analysts adopt in earnings conference calls. I argue that these patterns have a common origin in a reduced skepticism towards male CEOs. To make this case, I first document that analysts' beliefs about firm performance systematically under-react to bad news from male-led companies relative to the rational expectations benchmark, whereas they adjust their expectations rationally to similar bad news from female-led companies. Next, I show that investors also display this biased reaction to news, with stock returns reacting less negatively to negative surprises from male-led companies relative to their female-led peers on earnings announcement days. I then shed light on the underlying mechanism by constructing a text-based measure of disagreement from earnings conference calls. After negative surprises, analysts express less disagreement with the narrative conveyed by male-led firms relative to their female-led peers. This effect is entirely concentrated amongst male analysts, who represent, on average, more than 85% of the participants in these calls.

*I am grateful to my advisors, John Campbell, Katherine Coffman, Robin Greenwood, Samuel Hanson, and Tarek Hassan for their guidance and invaluable contribution. I also thank Fernando Cardoso, Elisabeth Kempf, Josh Lerner, Adi Sunderam, Ahmed Tahoun and several other seminar participants for their helpful comments.

†**Harvard University (email: mmello@hbs.edu)**

## 1. Introduction

In the earnings conference call for the third quarter of 2009, the CEO of MGIC Investment Corporation stated '*The third quarter was clearly a disappointing one financially as we reported a net loss of $517.8 million, with a diluted loss per share of $4.17.*' While financial markets should react to this news similarly regardless of whether the firm's CEO is named Jane or John, in this paper, I show that male CEOs, like John, tend to receive the benefit of the doubt. I document that financial markets react less to bad news about firm performance from male-led companies relative to their female-led peers and that this pattern reflects deviations from the rational expectations benchmark. This is surprising given the high-stakes nature of financial markets and the fact that there has been a thorough vetting process for the position of CEO.[1,2]

I document the asymmetric pattern of reaction to news by CEO gender across three independent settings and databases, which strengthens my confidence in this key finding. I first explore how analysts' beliefs about the earnings per share evolve over time, which allows me to measure deviations from the rational expectations benchmark. Next, I examine stock price movements around earnings announcements to determine whether investors display the same biases as analysts. Finally, I analyze the language used by executives and analysts during earnings conference calls, which sheds light on the mechanism driving my results.

Key to documenting these gender-based asymmetries in financial markets' reaction to news in my first setting is to measure both analysts' beliefs and new information about firm performance. Following Bouchad, Krüger, Landier, and Thesmar (2019), I measure beliefs using analysts' first forecasts after a given announcement date. Suppose I am interested in beliefs about firm performance for the third quarter of 2020 (Q3-2020). I collect analysts' forecasts for Q3-2020 made in two moments in time: first, right after the announcement of results for Q1-2020, and then again right after the announcement of results for Q2-2020.

[1] Any type of bias is costly when agents price public companies because influential market players profit either directly or indirectly from forming accurate expectations about future firm performance. Consider, for example, investors who must buy or sell company stocks, or analysts that provide client guidance about different companies.

[2] For instance, recent evidence suggests that traits that may affect managerial decisions — such as risk-taking — are similar across male and female candidates for CEO positions (Kaplan and Sorensen 2021).

Beliefs can change, so I measure new information relevant to investors' beliefs using analysts' forecast revisions, and define good news to be positive forecast revisions and bad news to be negative forecast revisions.[3]

I show that analysts' beliefs react asymmetrically to new information about firm performance based on CEO gender using predictive regressions of future forecast errors on past forecast revisions. These regressions provide a convenient way of measuring the sensitivity of beliefs to news because, under rational expectations, forecast errors should not be predictable (Coibion and Gorodnichenko 2015). A positive coefficient in these regressions points to an under-reaction to news relative to the rational expectations benchmark, while a negative coefficient points to an over-reaction.

I find that analysts react less to news from male-led companies relative to those from their female-led peers. This weaker reaction is driven by an under-reaction to news relative to the rational expectations benchmark for male-led companies: the estimated reaction coefficient for male-led companies is positive and statistically different from zero. For female-led companies there is neither under- nor over-reaction relative to the rational expectations benchmark, with an estimated reaction coefficient that is not significantly different from zero.

My first key result is that the gender-based asymmetry in analysts' reaction to news is state-dependent: the weaker reaction to new information from male-led companies relative to female-led firms occurs after the receipt of bad news. The state-dependency in reactions is driven by analysts' forecasts for male-led companies: after receiving bad news, beliefs about male-led companies under-react relative to the rational expectations benchmark. In contrast, after receiving good news, the coefficient on forecast revisions is not statistically different from zero. For female-led companies, after either good or bad news, no under- or over-reaction is detectable in the data. These results are not explained by other firm characteristics.

In words, this suggests that male CEOs — but not their female peers — receive the

[3]Note that I am implicitly assuming that beliefs change only because new information comes in, as is the case if agents are Bayesian, and ruling out behavioral belief changes caused by changes in mood or the influence of peers in the absence of information. Additionally, note that I am not assuming full rationality — Bayesian agents can be irrational in the sense of having unusual or extreme priors.

benefit of the doubt from analysts after bad news.

I use a simple model of Bayesian belief updating to interpret this key finding. In this simple framework, I assume analysts are trying to learn the firm's fundamental value, $v$, by observing earnings, a noisy signal of $v$. Moreover, there is a good (bad) state of the world with high (low) mean fundamental value. I also allow for one behavioral component in this otherwise standard model: agents can under- or over-estimate the precision of earnings as a signal of fundamental value. In this case, the coefficient on forecast revisions in the Coibion and Gorodnichenko (2015) framework is directly linked to the difference between the standard deviation of the noise under the agents' perceived and the objective data generating process. The intuition is that analysts react less to bad news from male-led companies relative to that from their female-led peers because they believe the precision of earnings as a signal of fundamental value in the bad state of the world is lower for these male-led companies.

Next, in my second setting, I show that this pattern of reaction to news is not exclusive to analysts, rather it is also displayed by investors. To that effect, I focus on earnings announcement days and measure market surprises based on analysts' last forecasts before the announcement day. I show that a pattern of weaker reaction to bad news from male- relative to female-led companies is also apparent in stock returns: after negative market surprises, abnormal returns associated with female-led companies are more than 64% more negative than those associated with male-led companies.

In my third and final setting — earnings conference calls — I again document evidence consistent with the gender-based asymmetries in analysts' reactions to news, and further shed light on the underlying mechanism. Earnings conference calls provide an ideal setting to explore these gender-based dynamics, as they bring together analysts — whose beliefs were analyzed in my first setting — and investors — whose behavior was linked to stock returns in my second setting — allowing both groups to engage directly with firm executives.

Based on the transcripts of these earnings conference calls, I develop a measure of semantic disagreement between analysts and executives determined by the distance between their sentiment scores. This measure captures disagreement in two ways. First, analysts may express disagreement by raising questions on more "negative topics" (i.e. topics with a more negative tone) than those addressed in the executives' presentations — that is, the firm's narrative. Second, they may disagree by using a more negative tone when discussing the same topics as the firm's narrative.

I find that, after negative market surprises, analysts express less disagreement with the narrative conveyed by male-led rather than by female-led companies. Using the transcripts from earnings conference calls also allows me to examine the gender of the analysts asking questions and reveals that the asymmetry in disagreement by CEO gender is driven solely by male analysts. For female analysts, there are no detectable differences in disagreement comparing calls for female- versus male-led companies. Importantly, male analysts represent, on average, more than 85% of the analyst pool in these earnings calls.

These findings suggest that male-led companies face less pushback from analysts as they attempt to convey bad news. In other words, analysts tend to be less skeptical or more credulous of male-led companies in these situations. This reduced skepticism toward male CEOs allows them to more easily set the narrative about the bad news for market participants. This credulity is reflected in the biased pattern observed in analysts' forecasts, stock prices, and disagreement.

**Literature Review.** The methodology in this paper builds on three main strands of literature. First, it builds on the work on biased belief updating. This literature has used predictive regressions of forecast errors on forecast revisions with macroeconomic variables and analysts forecasts to explore forecasters' biases, such as under-reaction to news (Coibion and Gorodnichenko 2015, Bordalo, Gennaioli, La Porta, and Shleifer 2019, Bouchad, Krüger, Landier, and Thesmar 2019, Bordalo, Gennaioli, Ma, and Shleifer 2020, Bordalo, Gennaioli, La Porta, and Shleifer 2022, Bordalo, Gennaioli, and Shleifer 2022, Kelly, Malamud, Siriwardane, and Wu 2024, Farmer, Nakamura, and Steinsson 2024). I extend the framework to explore the correlation in analysts' biases with CEO gender and sign of news (good versus bad news).

Second, this paper builds on a literature that has used event studies around earnings announcements to show that investors under-react to news on the announcement day. Some examples are Ball and Brown (1968), Foster, Olsen, and Shevlin (1984), Bernard and Thomas (1989), Bernard and Thomas (1990), Chan, Jegadeesh, and Lakonishok (1996), Jegadeesh and Titman (1993), Kothari and Warner (2007), and Hirshleifer, Lim, and Teoh (2009). I contribute by showing that the pattern of investors' reaction to news correlates with CEO gender.

Finally, this paper builds on research using text as data (e.g. Gentzkow, Kelly, and Taddy 2019, Bybee, Kelly, Manela, and Xiu 2021, Flynn and Sastry 2022, Kalyani, Bloom, Carvalho, Hassan, Lerner, and Tahoun *Forthcoming*). In particular, it builds on the literature that has used the text of corporate disclosures, news, or earnings conference calls to characterize shocks or explain stock returns (e.g. Tetlock, Saar-Tsechansky, and Macskassy 2008, Garcia 2013, Jiang, Lee, Martin, and Zhou 2019, Hassan, Hollander, Lent, and Tahoun 2019, Hassan, Hollander, Lent, and Tahoun 2024, Hassan, Schreger, Schwedeler, and Tahoun *Forthcoming*). I suggest a novel measure of semantic disagreement between analysts and executives and investigate its state-dependency.

This paper also contributes to the emerging literature on learning about others based on their gender.[4] The closest study in the literature is Sarsons (2024). The author uses Medicare data to understand how surgeons are evaluated for their surgical outcomes ("performance") in terms of medical referrals of new patients. The paper finds that female surgeons are more penalized for the death of a patient in surgery ("bad news") through lower referrals of new patients relative to male surgeons, but are treated similarly as their male peers after the unexpected survival of a patient ("good news"). I build on this important contribution in two ways. First, I show that a similar gender-based asymmetry occurs when investors evaluate the performance of female- and of male-led public companies. Second, because I directly observe individuals' beliefs — i.e. their forecasts — instead of only the actions associated with those beliefs — such as medical referrals —, I can compare beliefs to the rational expectations benchmark.

This paper is also related to a literature on the double standard of "punishment" based on gender. In the financial advisory industry, Egan, Matvos, and Seru (2022) show that, after an incident of misconduct, female advisors are more likely to lose their jobs and less likely

[4] One example is Sarsons, Gërxhani, Reuben, and Schram (2021), who find that there is a gender-based in difference in how credit for group work is attributed: conditional on quality and other observable variables, female academic economists — but not male ones — are less likely to receive tenure the more they coauthor. In another example, Coffman, Flikkema, and Shurchkov (2021) use an experimental setting to show that, conditional on the quality of ideas, individuals are less likely to be selected as team leader in gender incongruent domains.

to find new jobs relative to their male peers. Using information on executives' compensation and measuring firm performance through changes in a firm's market value, Albanesi, Olivetti, and Prados (2015) show that female executives are more exposed to bad performance and less exposed to good performance relative to their male peers. Finally, Landsman (2019) shows that female executives are more likely than their male peers to leave their firm following exogenous industry-wide contractions. I contribute by documenting a difference by CEO gender in the punishment for bad earnings news.

This literature has interpreted this double standard of punishment as suggestive of "in-group" bias,[5] and of female executives being less entrenched than male executives. Note that according to these explanations, the double standard in punishment can be interpreted as there being different thresholds by gender below which there is punishment. I contribute with a new interpretation. For instance, suppose that, in the current context, female CEOs would be punished if the perceived firm's fundamental value is below a certain threshold. Previous literature would argue that this threshold — and not necessarily perceived fundamental value — is different by CEO gender. The evidence I document in this paper suggests a different angle to this story: perceived fundamental value is different between female- and male-led companies.

This paper also contributes to an empirical literature on gender-based differences in the beliefs of investors. Lee and James (2007) find that stock markets react more negatively to the appointment of new female CEOs than to that of new male CEOs. Using data up to 2004, Wolfers (2006) finds no systematic differences in returns to holding stock in female-led firms relative to their male-led counterparts within S&P 1500 firms. The author, however, emphasizes that this result reflects the weak power of the tests performed given the small number of female CEOs in the sample. Niessen-Ruenzi and Ruenzi (2019) find significantly lower inflows in female-managed funds relative to male-managed ones, although there is no evidence of differences in performance. Finally, and closely related to this paper, using analysts' annual forecasts for earnings per share, Jannati, Kumar, Niessen-Ruenzi, and Wolfers (2023) find that male analysts have larger forecast errors (as measured by the

[5] By "in-group" bias, I mean a pattern of favoritism towards "in-group" members (e.g. men being more lenient when punishing men) over "out-group" members (e.g. men being comparatively less lenient when punishing women).

difference between the realization and the forecast) and worse stock recommendations for firms headed by female CEOs. These findings suggest that there is an average difference in analysts' beliefs about male- and female-led companies, with analysts being more pessimistic, on average, about female-led companies. My main contributions are showing that there is a difference in how investors learn (i.e. how they update their beliefs) about firm performance over time, and that this process is state-dependent.

Finally, while establishing the causal link between CEO gender and the reaction of financial markets to news about firm performance is a difficult problem, I make progress by documenting that investors' reaction to news is asymmetric according to CEO gender and that this pattern is not explained by other observable characteristics.

In section 2, I present the data and key measures for my first setting. In section 3, I explore analyst-level forecasts to document differences in analysts' reaction to news by CEO gender. In section 4, I show that this pattern of reaction documented for analysts spills over to the market and affects stock returns on earnings announcement days. Finally, in section 5, I document similar patterns of reaction using text data from earnings conference calls. Section 6 concludes.

## 2. Data and key measures

In this section, I review the data and key measures that I use to obtain my first main result. Aiming to investigate how investors' beliefs react to news about firm performance depending on CEO gender, I collect information on beliefs about firm performance and on objective firm performance — realizations of earnings per share — based on data for the universe of US public companies available at Institutional Brokers Estimate System (IBES). The sample starts at the second quarter of 1984 and ends at the fourth quarter of 2022. To obtain information on CEO gender, I combine the main dataset with Execucomp and WRDS Professional.

### 2.1. Data on CEO and firm characteristics

I collect information on CEO gender as well as CEO tenure in a firm, CEO age, and a firm's market capitalization. The reason for collecting additional information beyond CEO gender is to ensure that main results are not driven by observable differences between male- and female-led companies.

I am interested in determining the gender of top executives for the universe of US firms in IBES — that is, the set of firms for which I have data on analysts' performance forecasts. In order to do so, I use the following procedure. For firms available in Execucomp (firms in the S&P 1,500), gender is available in the data. For the remaining firms, I infer gender based on an algorithm using executives' names as recorded in WRDS Professional. From a total of 651,519 firm-CEO pairs available in these databases, only 72,865 (11% of the total) are dropped due to lack of information on CEO gender. Appendix C details this process.

Figure 1 shows the yearly evolution of the share of female-led companies — i.e., those with a female CEO or co-CEO — in the main sample since 1990. This share has been growing steadily over time, with the share of female-led companies around 6% as of 2022.

With respect to other CEO characteristics, I construct executive age and tenure at a particular company using data from Execucomp and from WRDS Professional. For firm characteristics, I use CRSP data to calculate market capitalization as a measure of firm size. See Appendix Table A1 for summary statistics of female- and male-led companies.

Throughout the paper, I refer to a firm-CEO pair, $i$, as a firm to facilitate the exposition. For most firms, this is a natural simplification since there is only one CEO at a given time. Some companies, however, have more than one CEO — that is, some co-CEOs — at the same time. In these cases, I run regressions at the firm-CEO level — so that a firm with two co-CEOs in a given period has two observations in the sample, one for each co-CEO. To ensure that firms with more co-CEOs are not overweighted in the sample, observations in all regressions are weighted by the inverse of the total number of CEOs in a company at a given period.

### 2.2. Forecast errors and forecast revisions

In order to show how beliefs react to news about firm performance, I use predictive regressions of future forecast errors on past forecast revisions (Coibion and Gorodnichenko 2015). I obtain these key variables using data from the Institutional Brokers Estimate System (IBES).

Specifically, I use information on forecasts and realizations of earnings per share from IBES unadjusted files. To ensure that these observations are comparable over time, I adjust them using CRSP adjustment factors.[6]

I assume that the forecast made at quarter $t$ contains all the information up to — and inclusive of — period $t$, as Figure 2 depicts. In particular, let $t$ be a given announcement date, $i$ a given company, and $a$ a given analyst. Throughout the paper, I let $X_{i,t}$ be the realization of company $i$'s earnings per share (EPS) announced at quarter $t$, and $\hat{E}_{a,t-1}(X_{i,t})$ be the forecast made by analyst $a$ at quarter $t-1$ for the company $i$'s EPS in quarter $t$. Then, the forecast error of analyst $a$ at period $t+1$ is

$$\text{Forecast Error}_{a,i,t+1} = \frac{X_{i,t+1} - \hat{E}_{a,t}(X_{i,t+1})}{P_{t-1}}, \tag{1}$$

and her forecast revision at period $t$ is

$$\text{Revision}_{a,i,t} = \frac{\hat{E}_{a,t}(X_{i,t+1}) - \hat{E}_{a,t-1}(X_{i,t+1})}{P_{t-1}}. \tag{2}$$

In particular, I consider the forecast $\hat{E}_{a,t}(X_{t+1})$ to be the first forecast posted by analyst $a$ for quarter $t+1$ in a short window of 28 days after the announcement day of quarter $t$. Imposing this limit ensures that forecast revisions capture solely the effect of the new information obtained on announcement day. I also require that a given announcement is at least 50 days apart from the next announcement day.

Aiming to avoid using stale forecasts, I only consider new forecasts made during the

[6] IBES also includes data on realizations and forecasts of earnings per share on an adjusted basis — i.e., adjusted for the effects of stock splits and stock dividends on per share amounts over time. However, IBES provides this adjusted data rounded to the nearest penny, which can cause loss of information (see Payne and Thomas 2003).

relevant 28-day period. The IBES file with unadjusted forecasts contains two time markers. One is the date when a forecast was first posted on the database and is the time variable I consider when building the database. The second is the date when IBES confirms with the analyst that a certain forecast is still valid.

I normalize forecast errors and revisions by the stock price to ensure comparability in the cross-section of companies. In particular, I use the end-of-period price of a stock two quarters before the earnings announcement date — thus ensuring that the price is predetermined relative to forecast errors and revisions.

An analysis of the frequency of analysts' revisions after an announcement day shows no economically sizeable differences by CEO gender — at least within company-announcement date pairs included in the final sample. First, I investigate the number of days that analysts take to post a new forecast after an announcement date. Appendix Table A2 shows that it takes, on average, 12 days ($p < 0.001$) for an analyst to post a new forecast for a male-led company. The excess number of days for female-led companies is -0.321 days ($p = 0.299$). Another way to look at the frequency of revisions is to evaluate the number of revisions that are made between two announcement dates. Analysts make, on average, 1.468 revisions ($p < 0.001$) for a male-led company. The excess number of revisions for female-led companies is -0.046 ($p = 0.029$).

The final dataset is obtained by requiring that forecast errors ($\text{Forecast Error}_{a,i,t+1}$), forecast revisions ($\text{Revision}_{a,i,t}$), and past earnings realizations ($X_{i,t-1}$) are not missing. Moreover, I require that past prices (close price measured two quarter before the announcement date), $P_{t-1}$, be larger than \$1. Forecast revisions are winsorized at the 1st and 99th percentiles. The final dataset contains data from the second quarter of 1984 until the fourth quarter of 2022.

## 3. Analysts' reaction to news about firm performance

In this section, I show that analysts react to news about firm performance asymmetrically conditional on the gender of the CEO leading the company. To that effect, I document that analysts' forecast revisions for male-led companies — but not for female-led ones — predict their forecast errors. This predictability is concentrated in bad states of the world — that is, after receiving bad news.

### 3.1. Measuring analysts' reaction to news

I use the methodology introduced in the seminal work of Coibion and Gorodnichenko (2015): that is, I run predictive regressions of forecast errors on forecast revisions. While many papers have used this framework in different contexts (e.g. Bordalo, Gennaioli, Ma, and Shleifer 2020, Bordalo, Gennaioli, La Porta, and Shleifer 2022, Bordalo, Gennaioli, and Shleifer 2022, among others), fewer studies have used it to analyze analysts' forecasts of the earnings per share. Bordalo, Gennaioli, La Porta, and Shleifer (2019) show that analysts' long-run forecasts about earnings growth over-react to news relative to the rational expectations benchmark. Kelly, Malamud, Siriwardane, and Wu (2024) show that analysts 6-quarters-ahead forecasts for the earnings-per-share over-react to news relative to the rational expectations benchmark. Meanwhile, Bouchad, Krüger, Landier, and Thesmar (2019) show that analysts shorter-run forecasts — their annual forecasts — under-react relative to the rational expectations benchmark.[7]

Predictive regressions of forecast errors on forecast revisions provide a convenient method for studying belief updating because they are informative about the responsiveness of beliefs to new information. As noted in Coibion and Gorodnichenko (2015), forecast revisions at a given period $t$ summarize all the information received by the forecaster at that period and thus, under rational expectations, must be uncorrelated with forecast errors in $t+1$.

Specifically, the coefficient ($\beta$) of a regression of forecast errors on forecast revisions,

$$\text{Forecast Error}_{t+1} = \alpha + \beta \times \text{Revision}_t + e_t, \tag{3}$$

should equal zero under rational expectations. In contrast, a coefficient $\beta > 0$ indicates under-reaction in agents' expectations: between $t-1$ and $t$, agents received some piece of

[7]The observed discrepancy — with analysts' shorter-run expectations displaying under-reaction, but their longer-run expectations displaying over-reaction — is consistent with Augenblick, Lazarus, and Thaler (*Forthcoming*).

news and revised expectations accordingly in the correct direction, say a positive (negative) revision, but they did so by systematically less than they should have, thus incurring in positive (negative) forecast errors. A coefficient $\beta < 0$ is indicative of the opposite: agents' expectations over-react to new information. Bouchad, Krüger, Landier, and Thesmar (2019) use this method to study under-reaction of analysts' expectations to firm profitability. Bordalo, Gennaioli, La Porta, and Shleifer (2019) use it to diagnose over-reaction in analysts' long-run growth expectations.

Note that under the null hypothesis of rational expectations, agents are assumed to follow the Bayesian updating rule. If agents are Bayesian, they should only revise their expectations when there is new information. Hence, under the null of rational expectations, forecast revisions should solely reflect the arrival of news. However, if agents deviate from Bayesian updating, interpreting forecast revisions can be trickier. In this paper, I continue to interpret forecast revisions as indicative of new information, even if rational expectations fails. When agents do not follow Bayesian updating and revise their forecasts without new information being released — say, because of a change in their mood about the future —, the interpretation of regressions of forecast errors on forecast revisions changes. Instead of viewing these regressions as revealing how agents react to news, we should interpret them as reflecting agents' reactions to changes in their mood about the future.

I first estimate the regression in (3) separately by CEO gender — i.e. in sub-samples of only male- and female-led companies. For the sub-sample of company-period observations that are female-led (male-led), I run the regression

$$\text{Forecast Error}_{a,i,t+1} = b_0 + b_1 \times \text{Revision}_{a,i,t} + u_{a,i,t+1}. \tag{4}$$

The sub-sample of female-led companies is determined as follows: I consider company $i$ at quarter $t$ to be female-led if it has a female CEO or co-CEO during that period.

Panel A of Table 1 displays the results by sub-sample: while I detect under-reaction relative to the rational expectations benchmark for male-led companies, I find neither under-nor over-reaction for female-led companies.[8] Columns 1 and 3 present results without adding any additional controls. Columns 2 and 4 control for past earnings and firm size, and for a series of fixed effects (forecast period, analyst, broker, and firm fixed effects). Importantly, these results are also not explained by differences in firm performance between male- and female-led firms.[9]

In my preferred specification for male-led companies, a 1 p.p. increase in forecast revisions predicts a 0.753 p.p. (p = 0.016) increase in forecast errors. For female-led companies, the equivalent estimated coefficient is 0.069 (p = 0.647) and is not statistically different from zero. Note that the magnitude of the coefficient for male-led companies is in line with the findings in previous literature. For example, using annual data on analysts' earnings forecasts, Bouchad, Krüger, Landier, and Thesmar (2019) document an under-reaction coefficient of between 0.165 and 0.176. In turn, Coibion and Gorodnichenko (2015) document under-reaction coefficients between 1.062 and 1.196 using quarterly forecasts for macroeconomic variables.

The second exercise I perform to diagnose whether analysts' reaction to news differs by CEO gender is to estimate an extended version of (4) with the full sample of firms. In this case, let $1(\text{Female-led})_{i,t}$ be an indicator function that takes the value of 1 if firm $i$ in period $t$ has a female CEO or co-CEO — that is, if firm $i$ in period $t$ is female-led. I am interested

[8] The data also has evidence of another type of mistake committed by analysts: average forecast errors by CEO gender do not necessarily average zero. Specifically, estimates are suggestive of analysts being more optimistic about male-led companies relative to their female-led peers. This result, detailed in Appendix F, is consistent with the findings in Jannati, Kumar, Niessen-Ruenzi, and Wolfers (2023) with analysts' annual forecasts. This type of mistake cannot account for the results in this section, which are about the sensitivity of beliefs to news rather than the average level of those beliefs.

[9] See Appendix Table A3 for an analysis of various moments of firm performance, including average, volatility, and persistence, across three performance measures: earnings per share, and its quarterly and yearly growth rates. When testing for performance differences between male- and female-led companies across these nine dimensions, using a 5% significance level with a Bonferroni adjustment, we fail to reject the hypothesis of no differences between these companies. Appendix D provides further details on this analysis. Importantly, any differences in objective firm performance should not affect my results, which are about deviations from the rational expectations benchmark. As long as analysts correctly perceive potential differences in objective performance between male- and female-led firms, no over- or under-reaction should occur. Appendix E presents a model where analysts under-estimate the persistence of firm performance of male-led companies, but the data rejects its key testable hypothesis.

in the regression:

$$\text{Forecast Error}_{a,i,t+1} = \beta_0 + \beta_1 \times 1(\text{Female-led})_{i,t} + \\ \beta_m \times \text{Revision}_{a,i,t} + \beta_f \times \text{Revision}_{a,i,t} \times 1(\text{Female-led})_{i,t} + u_{a,i,t+1}. \tag{5}$$

The results for the stacked specification (5), in panel B of Table 1, confirm the findings by sub-sample. Since there are relatively few female CEOs in the sample, the under-reaction observed in male-led companies holds on average across the full sample, aligning with the short-term under-reaction to annual forecasts documented by Bouchad, Krüger, Landier, and Thesmar (2019).

### 3.2. State-dependency of analysts' reaction to news

To gain a better understanding of the key result in the last section — that analysts react less to news from male-led companies relative to that from their female-led peers —, I explore the effects of different types of news — good or bad.

I start by defining good news as positive forecast revisions and bad news as negative forecast revisions. Measuring news through forecast revisions is a natural choice in the framework of Coibion and Gorodnichenko (2015) because forecast revisions summarize all the information received by the forecaster at a given period. One way to think about this choice is that a forecast revision captures the forecaster's own interpretation about how some underlying piece of information impacts earnings.

I estimate equation (5) presented in the last section by sub-sample of good news or bad news. Table 2 presents the results: panel A contains estimates for the sub-sample of negative forecast revisions — bad news — and panel B contains estimates for the sub-sample of positive forecast revisions — good news. Column 1 in both panels present baseline estimates, while column 2 introduces past earnings and firm size as controls, and column 3 further controls for a series of fixed effects (forecast period, analyst, firm, and broker fixed effects).[10]

[10] I document that the findings in Table 1 — at the analyst-level — also hold at the consensus-level. To that effect, I first aggregate the forecast errors and forecast revisions by taking the average across all analysts, for each period and firm. I then re-estimate the regression specification in (5) for the sub-sample of positive and of negative consensus revisions. See Appendix Table A8 for results.

The estimates by sub-sample of good and bad news suggest that the weaker reaction to news from male-led companies relative to their female-led peers is concentrated after the receipt of bad news.[11] Indeed, the estimates in panel A for $\beta_f$ (see equation (5)) — which captures the difference in reaction between female-led and male-led companies — after negative forecast revisions are negative and significant. In contrast, the estimates for this same coefficient in panel B, after positive forecast revisions, are positive and not statistically different from zero.

The driver for this asymmetry is an under-reaction to bad news from male-led companies relative to the rational expectations benchmark.[12,13] The reaction coefficient $\beta_m$ (see equation (5)) for male-led companies after bad news is positive and statistically significant at least at the 10% level. In contrast, after good news, the estimates for the reaction coefficient $\beta_m$ are not statistically different from zero. In my preferred specification, $\beta_m$ is estimated at 1.072 (p = 0.025) after bad news and at -0.020 (p = 0.916) after good news. For female-led companies there is neither under- nor over-reaction after either good or bad news. In my preferred specification, the reaction coefficient for female-led companies after bad news — obtained by summing the estimates for $\beta_m$ and $\beta_f$ — is 0.021 (p = 0.931), and, after good news, 0.032 (p = 0.852).

To check whether the differences in reaction by sign of news are significant, I consider an

[11]Importantly, this key result is documented even if we consider more recent samples, from 2000 to 2022 or from 2015 to 2022, as Appendix Table A10 shows.

[12]Note that if we condition only on the most recent sample, from 2015 to 2022, the driver for this asymmetry by CEO gender is an over-reaction to bad news from female-led companies — see Appendix Table A10 for details.

[13]This asymmetry in the reaction of analysts by "state of the world" — whether it's good news or bad news being observed — is in line with the findings in Kelly, Malamud, Siriwardane, and Wu (2024).

extended version of equation (5),

$$
\begin{aligned}
\text{Forecast Error}_{a,i,t+1} = {} & B_0 + B_1 \times 1(\text{Female-led})_{i,t} + B_2 \times 1(\text{Revision}_{a,i,t} < 0) + \\
& B_3 \times 1(\text{Revision}_{a,i,t} < 0) \times 1(\text{Female-led})_{i,t} + \\
& B_{m,g} \times \text{Revision}_{a,i,t} \times 1(\text{Revision}_{a,i,t} \geq 0) + \\
& B_{f,g} \times \text{Revision}_{a,i,t} \times 1(\text{Revision}_{a,i,t} \geq 0) \times 1(\text{Female-led})_{i,t} + \\
& B_{m,b} \times \text{Revision}_{a,i,t} \times 1(\text{Revision}_{a,i,t} < 0) + \\
& B_{f,b} \times \text{Revision}_{a,i,t} \times 1(\text{Revision}_{a,i,t} < 0) \times 1(\text{Female-led})_{i,t} + \\
& u_{a,i,t+1}.
\end{aligned}
\tag{6}
$$

Table 3 summarizes the results of this fully stacked specification and confirms that the reaction coefficients $B_{m,g}$ and $B_{m,b}$ in equation (6) are statistically different from one another.

### 3.3. Interpretation: a simple model of the "benefit of the doubt"

I use a simple model of Bayesian belief updating to understand the results in the previous section. Such model formalizes the finding that male-led companies receive the benefit of the doubt after bad news. Specifically, in the model, agents under-estimate the precision of earnings as a signal of fundamental firm performance for male-led companies in bad states of the world.

Consider a simple framework in which analysts are trying to learn the firm's fundamental value, $v$, but only observe earnings, $s_t$, a noisy signal of this fundamental value. In this context, suppose that there is a good state of the world, $\theta_{good}$, with high fundamental value — that is $v|\theta_{good} \sim N(\mu_{good}, \sigma_v^2)$ —, and a bad state of the world, $\theta_{bad}$, with low fundamental value — $v|\theta_{bad} \sim N(\mu_{bad}, \sigma_v^2)$ with $\mu_{good} > \mu_{bad}$. Earnings, the noisy signal of firm fundamental value, evolves according to $s_t = v + \eta_t$, where $\eta_t \sim_{iid} N(0, \sigma_\eta^2)$ and $\eta_t$ is drawn independently of the state of the world $\theta$. Finally, assume that the bad state of the world occurs with probability $p \in [0, 1]$ and, therefore, the good state of the world occurs with probability $1 - p$.

For simplicity, I assume that agents hold correct prior beliefs about the firm's fundamental value conditional on the state of the world. They, however, hold potentially biased beliefs about the distribution of noise: depending on the type of company and the state of the world, agents may under- or over-estimate the precision of earnings as a signal of fundamental value. This assumption means that, under the agents' perceived data generating process, $\eta_t|\theta \sim N(0, \hat{\sigma}^2_{\eta,\theta})$.

At period $t = 0$, nature draws the state of the world. Agents, however, only learn about this state in period $t = 1$ when they observe the first earnings result, $s_1$. So, in period $t = 0$, agents form forecasts for earnings in periods $t = 1$ and $t = 2$ according to their prior beliefs. In period $t = 1$, they observe both the state of the world, $\theta$, and the earnings realization, $s_1$, and revise their forecast for earnings in period $t = 2$.

I am interested in understanding the Coibion and Gorodnichenko (2015) regression, summarized in equation in (3), in this context. To that effect, note that, under the previous assumptions, the forecast revision for male-led companies in $t = 1$ is

$$
\begin{aligned}
fr_1 &= \underbrace{\hat{E}(s_2|s_1)}_{posterior} - \underbrace{\hat{E}(s_2)}_{prior} \\
&= 1(\theta = \theta_{good})\left[\mu_{good} + \frac{\sigma_v^2}{\sigma_v^2 + \hat{\sigma}^2_{\eta,\theta_{good}}}(s_1 - \mu_{good})\right] + \\
&\quad 1(\theta = \theta_{bad})\left[\mu_{bad} + \frac{\sigma_v^2}{\sigma_v^2 + \hat{\sigma}^2_{\eta,\theta_{bad}}}(s_1 - \mu_{bad})\right] - \\
&\quad [p\mu_{bad} + (1-p)\mu_{good}],
\end{aligned} \tag{7}
$$

and the forecast error in $t = 2$ is

$$
\begin{aligned}
fe_2 &= s_2 - \hat{E}(s_2|s_1) \\
&= 1(\theta = \theta_{good})\left[\frac{\hat{\sigma}^2_{\eta,\theta_{good}}}{\sigma_v^2 + \hat{\sigma}^2_{\eta,\theta_{good}}}(v - \mu_{good}) + \eta_2 - \frac{\sigma_v^2}{\sigma_v^2 + \hat{\sigma}^2_{\eta,\theta_{good}}}\eta_1\right] + \\
&\quad 1(\theta = \theta_{bad})\left[\frac{\hat{\sigma}^2_{\eta,\theta_{bad}}}{\sigma_v^2 + \hat{\sigma}^2_{\eta,\theta_{bad}}}(v - \mu_{bad}) + \eta_2 - \frac{\sigma_v^2}{\sigma_v^2 + \hat{\sigma}^2_{\eta,\theta_{bad}}}\eta_1\right].
\end{aligned} \tag{8}
$$

From (3), the reaction coefficient $\beta$ conditional on the state of the world is determined by

$$\beta_\theta = \frac{Cov(fe_2, fr_1|\theta)}{Var(fr_1|\theta)} = \frac{\hat{\sigma}^2_{\eta,\theta} - \sigma^2_\eta}{\sigma^2_v + \sigma^2_\eta}. \tag{9}$$

In words, equation (9) indicates that the reaction coefficient is directly linked to the difference between the standard deviation of prior beliefs about noise and that of the true distribution. In particular, there is under-reaction ($\beta_\theta > 0$) if agents under-estimate the precision of earnings as a signal of fundamental value in state $\theta$ ($\sigma^2_\eta < \hat{\sigma}^2_{\eta,\theta}$).

One interesting result from this simple model is that positive forecast revisions tend to be correlated with the good state of the world and negative forecast revisions, with the bad state of the world. To see why, let's compute the conditional probability of negative forecast revisions in each state of the world:

$$\begin{aligned}
\mathbb{P}(fr_1 < 0|\theta_{good}) &= \mathbb{P}\left[\underbrace{\frac{s_1 - \mu_{good}}{(\sigma^2_v + \sigma^2_\eta)^{1/2}}}_{\text{standard normal}} < \underbrace{-\frac{\sigma^2_v + \hat{\sigma}^2_{\eta,\theta_{good}}}{(\sigma^2_v + \sigma^2_\eta)^{1/2}\sigma^2_v} \cdot p \cdot (\mu_{good} - \mu_{bad})}_{<0} \middle| \theta_{good}\right] \\
\mathbb{P}(fr_1 < 0|\theta_{bad}) &= \mathbb{P}\left[\underbrace{\frac{s_1 - \mu_{good}}{(\sigma^2_v + \sigma^2_\eta)^{1/2}}}_{\text{standard normal}} < \underbrace{\frac{\sigma^2_v + \hat{\sigma}^2_{\eta,\theta_{bad}}}{(\sigma^2_v + \sigma^2_\eta)^{1/2}\sigma^2_v} \cdot (1-p) \cdot (\mu_{good} - \mu_{bad})}_{>0} \middle| \theta_{bad}\right]
\end{aligned} \tag{10}$$

The first expression shows that the conditional probability of negative forecast revisions in the good state of the world is smaller than 50%. This probability is decreasing in the difference between the mean fundamental value in the good and bad states of the world. The second expression shows that the conditional probability of negative forecast revisions in the bad state of the world is larger than 50%, and is increasing in the difference between the mean fundamental value in the good and bad states of the world. These probabilities determine, from an econometrician's point of view, when negative (positive) forecast revisions are a good proxy for the bad (good) state of the world.

Based on this simple model, the results of sections 3.1 and 3.2 allow us to back out one main bias from the data: for male-led companies, $\hat{\sigma}^2_{\eta,\theta_{bad}} > \sigma^2_\eta$. In words: the "benefit of the doubt" documented in previous sections can be formalized by assuming that agents believe that earnings results are poor signals of firm fundamental value for male-led companies in the bad state of the world. In contrast, for male-led companies in the good state of the world, $\hat{\sigma}^2_{\eta,\theta_{good}} = \sigma^2_\eta$, while for female-led companies $\hat{\sigma}^2_{\eta,\theta_{good}} = \hat{\sigma}^2_{\eta,\theta_{bad}} = \sigma^2_\eta$. A key intuition from these assumptions is that analysts react less to bad news from male-led companies relative to that from their female-led peers because they believe the precision of earnings as a signal of fundamental value in the bad state of the world is lower for these male-led companies.

Next, I parametrize the model for male-led companies and explore the conditions under which it can replicate the under-reaction observed in the data. To this end, the key moment that I match is the variance of the observed earnings-per-share (normalized by past prices), $Var(s_t) = 0.125$. Hence, I set $\sigma^2_v + \sigma^2_\eta = 0.125$. For simplicity, I assume that $\sigma^2_v = \sigma^2_\eta = \frac{0.125}{2}$. Changing $\sigma^2_v$ affects the overall under-reaction observed in the bad state of the world (see equation (9)): in particular, the larger is $\sigma^2_v$, then the smaller is the necessary difference $\hat{\sigma}^2_\eta - \sigma^2_\eta$ to match the under-reaction observed in the data after negative revisions. I also match the average earnings-per-share (normalized by past prices) in the data, $\mathbb{E}(s_t) = 0.072$. To that effect, I assume that the states of the world are equally as likely, so $p = 0.5$, and set $\mu_{bad} = 0$ and $\mu_{good} = \frac{0.072-0\cdot 0.5}{1-0.5} = 0.144$.

Figure 3 shows the resulting simulated reaction coefficient after negative and positive forecast revisions based on this parametrization. These estimates are based on 1,000 simulations with 1,000 observations each, and consider a parameter $\hat{\sigma}_{\theta_{bad}} = \phi \cdot \sigma_\eta$. The blue line in panel (a) marks the estimate in column 1 of panel A in Table 2 for male-led companies, and suggests that any $\phi \geq 2.5$ is enough to rationalize the exact coefficient observed in the data. Additionally, we see that any coefficient $\phi$ rationalizes the exact estimate for the reaction coefficient observed in the data after good news, as the blue line in panel (b) (which displays the estimate in column 1 of panel B in Table 2) is fully contained in the 90% confidence interval implied by the simulations.[14]

[14] As panel (b) in Figure 3 depicts, there is a slight over-reaction after positive forecast revisions. To understand why, note that this over-reaction is determined by the covariance $cov(fe_{t+1}, fr_t | fr_t \geq 0)$, which is defined as $\mathbb{E}(fe_{t+1} \times fr_t | fr_t \geq 0) - \mathbb{E}(fe_{t+1} | fr_t \geq 0) \times \mathbb{E}(fr_t | fr_t \geq 0)$. In the bad state of the world, when forecast errors are positive, they often follow a positive forecast revision. This makes the second part, $\mathbb{E}(fe_{t+1} | fr_t \geq 0) \times \mathbb{E}(fr_t | fr_t \geq 0)$, greater than zero. However, positive forecast revisions tend to happen in the good state of the world, where there is no over- or under-reaction. This makes the first part, $\mathbb{E}(fe_{t+1} \times fr_t | fr_t \geq 0)$, close to zero. As a result, the overall covariance ends up being slightly negative, mainly because of the influence of the second term.

In the baseline parametrization, with $\hat{\sigma}_\eta = 2.5 \cdot \sigma_\eta$ and $p = 0.5$, the estimated mean probabilities $\mathbb{P}(fr_t < 0|\theta_{bad})$ and $\mathbb{P}(fr_t < 0|\theta_{good})$ are, respectively, 0.930 and 0.340. We can then use Bayes' rule to show that, from an econometrician's point of view, negative forecast revisions proxy for the bad state of the world — with an average $\mathbb{P}(\theta_{bad}|fr_t < 0)$ of 0.732 — and positive forecast revisions proxy for the good state of the world — with an average $\mathbb{P}(\theta_{good}|fr_t \geq 0)$ of 0.904.

Based on this parametrization, I am not able to match the estimate for the unconditional reaction coefficient found in the data — of 0.753 (see Table 1) —, although the respective 95% confidence interval for this estimate — of [0.150, 1.357] — overlaps with the upper part of the simulation's confidence interval when $\hat{\sigma}_\eta \geq 2.5 \cdot \sigma_\eta$. For example, when I set $\hat{\sigma}_\eta = 2.5 \cdot \sigma_\eta$, the simulation mean is 0.141 with 90% confidence interval of [0.029, 0.253]. The key parameter in matching the unconditional reaction coefficient is the probability of the bad state of the world occurring, $p$. Appendix Figure B4 shows what happens with this coefficient as I vary $p$, but maintain the remaining parametrization constant (with $\hat{\sigma}_\eta = 2.5 \cdot \sigma_\eta$). Matching the exact estimate for the unconditional coefficient observed in the data would require setting $p$ at around 0.9.

Finally, note that the kind of asymmetry in the model presented in this section could be motivated from thinking through the lenses of the work on stereotypes and selective memory (e.g. Bordalo, Coffman, Gennaioli, and Shleifer 2016, Bordalo, Coffman, Gennaioli, and Shleifer 2019, Bordalo, Burro, Coffman, Gennaioli, and Shleifer *Forthcoming*). Motivated by the fact that most superstar CEOs are male — think about Jeff Bezos, Jamie Dimon or Tim Cook —, imagine that analysts have a positively skewed prior distribution of talent for male CEOs but not for female CEOs. In this case, analysts put little weight on the possibility of male CEOs being incompetent — thus treat bad news as noise —, while allowing for the possibility of these male CEOs being superstars. If analysts have a symmetric prior about the talent of female CEOs, then there would be a weaker reaction to bad news from male-led companies relative to their female-led peers.

### 3.4. Robustness

#### 3.4.1. Other measures of good versus bad news

I show that the documented gender-based asymmetries in reaction to new information by sign of news are robust to changing the underlying measure of news. While the baseline measure of good (bad) news is positive (negative) forecast revisions, here, I consider three alternative measures of news. First, I measure news using analyst-level past forecast errors, and define positive forecast errors as good news and negative forecast errors as bad news. Next, I consider two measures of news at the firm- rather than the analyst-level: the average forecast error across analysts, and the abnormal return of the company's stock over the S&P 500 registered on announcement day. In particular, I define a positive (negative) average forecast error or a positive (negative) abnormal return on announcement day as good (bad) news.

Table A4 summarizes results for (5), by sub-sample of good and bad news, for the alternative measures of good or bad news. As in the baseline, the difference in reaction between female-led companies and male-led ones, captured by $\beta_f$ (see equation (5)), is negative and significantly different from zero across the various measures of bad news. After good news, as was the case with the baseline, the estimated $\beta_f$ is smaller and, in most cases, not statistically different from zero.

In line with the previous findings, the estimates for $\beta_m$ (see equation (5)) under the alternative measures also point to an under-reaction to bad news from male-led companies relative to the rational expectations benchmark. The key difference with respect to the baseline results, however, is that estimates indicate some under-reaction too after good news. These estimated reaction coefficients after good news, however, are smaller than the respective ones after bad news. A stacked specification rejects the null that the estimated $\beta_m$'s after good and bad news are equal, with a p-value of 0.058 when the measure of news is the analyst-level forecast error (panel A). The p-value for the equivalent tests in panels B and C — with news measured by firm-level average forecast errors and abnormal return on announcement day — is, respectively, 0.098 and 0.062.

### 3.4.2. Changing the baseline specification

I show that baseline results are robust to changing the regression specification, as summarized in Table A5. Panel A shows robustness for the more general result that there is less reaction to news from male-led companies relative to their female-led peers (column 1, panel B, Table 1). Panel B shows robustness for the result that analysts react less to bad news from these male-led companies than they do to that from their female-led counterparts (column 1, panel A, Table 2). Panel C shows robustness for the result that there is no such asymmetry in reaction after the receipt of good news (column 1, panel A, Table 2). Column 1 replicates baseline results.

In the baseline regressions, observations are given the same weight across analysts which implies a form of value-weighting across companies because larger companies have more analysts covering them (see Appendix Figure B2). The first change I make is letting regressions be equally-weighted across firms. Column 2 across panels A through C in Table A5 summarizes results for these equal-weighted regressions, with only small differences in the estimated coefficients relative to the baseline.

The second change I make, in column 3 of Table A5, is to remove forecast revisions at the 1st and 99th-percentiles from the sample instead of winsorizing them at these levels — as I do in the baseline —, with the objective of understanding whether results are driven by these outliers. Results remain qualitatively the same. The key difference between columns 1 and 3 is a reduction in the documented under-reaction coefficient for male-led companies after bad news, which falls from 1.072 ($p = 0.025$) to 0.481 ($p = 0.026$). This decrease in the reaction coefficient, $\beta_m$ (see equation (5)), is related to the fact that the under-reaction to bad news from male-led companies relative to the rational expectations benchmark is concentrated, at the analyst level, on larger forecast revisions.[15]

[15]Appendix Table A6 depicts this fact: column 1 shows the baseline result with all negative forecast revisions, while column 2 shows results for small and negative forecast revisions (those above the median of the distribution), and column 3 shows results for large and negative forecast revisions (those below the median of the distribution). In particular, the reaction coefficient for male-led companies, $\beta_m$, is estimated at 1.072 ($p = 0.025$) in the baseline, at 0.275 ($p = 0.037$) for small revisions, and at 1.212 ($p = 0.026$) for large revisions. A stacked specification indicates that these coefficients — for small and large revisions — are statistically different from one another at the 5% level (difference estimated at 0.937 with a p-value of 0.039). If large forecast revisions occur in response to stronger signals — i.e., precise signals —, then one explanation for these results could come from the work of Augenblick, Lazarus, and Thaler (*Forthcoming*), who find under-reaction to strong signals (and over-reaction to weak signals).

The next robustness exercise I perform, contained in Appendix Table A8 and A9, is to re-estimate the baseline regressions with consensus-level forecasts instead of analyst-level forecasts. Specifically, I run regressions with both the mean and the median forecasts across analysts. The pattern of under-reaction to bad news from male-led companies but not to that from female-led companies is clear and more precisely estimated relative to that with analyst-level forecasts. This reinforces the relevance of the analyst-level results because consensus-level estimates tend to be more robust to idiosyncratic noise than analyst-level ones (Kučinskas and Peters 2022).

The last robustness exercise I perform is estimating quantile regressions — that is, regressions that target the different quantiles of forecast errors instead of their mean. Results are summarized in Appendix Figure B3, with panel (a) showing results after negative revisions and panel (b) showing results after positive revisions. In panel (a) we see that, in line with baseline results, analysts react *less* to bad news from male-led companies relative to that of their female-led peers for all quantiles of forecast errors. This is captured through a reaction coefficient $\beta$ that is more positive for male-led companies relative to female-led ones for all quantiles of forecast errors. The difference between reaction coefficients for male- and female-led companies is statistically significant at least at the 5% level for all quantiles after negative revisions. In contrast, after positive forecast revisions — again in line with baseline results — the reaction coefficients estimated for each quantile for male-led and female-led companies are not statistically different from one another.

### 3.4.3. Controlling for additional characteristics

I show that results are robust to adding further controls to the already heavily controlled specification in column 3 of panels A and B in Table 2. In this specification, I include a series of controls and fixed effects (analyst, forecast period, firm, and broker fixed effects). In particular, I control for: (i) the past level of an analyst's forecast, $\hat{E}_{a,t-1}(X_{i,t+1})$; (ii) firm size (market capitalization in $t-1$); (iii) an indicator of whether firm size is lower than \$2 billion, and (iv) its interaction with forecast revisions; (v) an indicator of whether firm size is higher than \$10 billion, and (vi) its interaction with forecast revisions. This means that neither fixed firm characteristics, nor fixed characteristics of analysts, brokers, time period, and firm size explain the differences in reaction coefficients by CEO gender.

I take this specification and further include an indicator of whether a certain observation is in the top 50% of the distribution of a given characteristic and its interaction with forecast revisions:

$$
\begin{aligned}
\text{Forecast Error}_{a,i,t+1} = \beta_0 + \beta_1 \times 1(\text{Female-led})_{i,t} + \\
&\beta_2 \times 1\{\text{Characteristic}_{i,t} \text{ in Top } 50\%\} + \\
&\beta_m \times \text{Revision}_{a,i,t} + \\
&\beta_f \times \text{Revision}_{a,i,t} \times 1(\text{Female-led})_{i,t} + \\
&\beta_{top} \times \text{Revision}_{a,i,t} \times 1\{\text{Characteristic}_{i,t} \text{ in Top } 50\%\} + \\
&u_{a,i,t+1}
\end{aligned} \tag{11}
$$

The idea is to test whether other CEO and firm characteristics explain the gender difference in reaction to news. I consider the following characteristics: CEO age, CEO tenure at the company, and a firm's past price, $P_{t-1}$.

Table A7 displays the results. Column 1 displays the estimates in column 3 of panels A and B in Table 2. Column 2 shows results when I further control for CEO tenure, column 3, for CEO age, and column 4, for past prices. The main takeaway is that controlling for these characteristics does not affect numerically or qualitatively the main results.[16]

## 4. Market's reaction to new information

In the previous section, I documented that analysts react asymmetrically to bad news about firm performance conditional on CEO gender. In this section, I show that this pattern is not restricted to analysts, with investors displaying the same bias. To that effect, I run regressions of stock market reaction on different measures of market surprise on earnings announcement days interacted with CEO gender. I find that stock markets react less negatively to negative surprises from male-led companies relative to that from their female-led counterparts. Such asymmetry is not present after positive surprises.

[16]Interestingly, a heterogeneity exercise suggests that the under-reaction to bad news from male-led companies is concentrated in cheaper companies (past prices below the median). Appendix Table A11 shows the results of this investigation.

### 4.1. Data

Following a vast literature on event studies in finance, I focus on earnings announcement days to understand whether market surprises affect a firm's stock price differently depending on CEO gender, and build standard measures of market surprise and market reaction for the universe of US public companies in IBES.

Market surprises are related to the amount by which a firm's earnings realization differs from market expectations. I measure the surprise observed by an analyst as the difference between the earnings realization and that analyst's last forecast before the announcement date. For example, suppose I am interested in measuring the surprise on the announcement day for the results of Q3-2020. I then collect an analyst's last forecast before that announcement day and subtract it from the earnings realization.

I build two measures of market surprise. The first one is the average sign of analyst-level surprises (called the fraction of misses on the same side or FOM score), following Chiang, Dai, Fan, Hong, and Tu (2019):

$$FOM = \frac{N_{+} - N_{-}}{N}, \tag{12}$$

where $N_{+}$ is the number of analysts that were positively surprised, $N_{-}$ is the number of analysts that were negatively surprised, and $N$ is the total number of analysts that made forecasts for a certain event. The second measure is the consensus surprise normalized by a firm's past stock price

$$\frac{CS}{PastPrices} = \frac{\frac{1}{N}\sum_{a} \text{Surprise}_{a}}{\text{Price}_{t-20days}}. \tag{13}$$

where $\text{Surprise}_{a} = \text{Realization} - \text{Last Forecast}_{a}$, and $a$ represents an analyst.

The final dataset contains information for these two measures from the first quarter of 1993 to the fourth quarter of 2022. I winsorize observations for $\frac{CS}{PastPrices}$ on the 1st and 99th percentiles. Because the FOM score varies between -1 and 1, I perform no further adjustments to that measure. I only include in regressions observations whose closing price measured 20 days before the announcement day, $\text{Price}_{t-20days}$, is higher than \$1, and require a minimum number of at least two analysts' forecasts for any given company. I consider only announcement dates for a given company that are at least 50 days apart.

Importantly, the average market surprise across the different measures is similar between companies led by male and female CEOs (see Appendix Table A14).

Market reaction is measured by the abnormal return of a firm's stock over the S&P 500 index on announcement day. Stock market prices are from CRSP, from where I also obtain returns on the S&P 500. The abnormal return around the announcement day is defined as

$$AR_{i,t} = r_{i,t} - r_{S\&P,t},$$

where $t$ is the announcement date, $i$ is a firm, $r_{i,t}$ is the stock's return on date $t$, and $r_{S\&P,t}$ is the return on the S&P 500 index on date $t$.

### 4.2. Market's reaction to market surprises

I use a traditional event study approach around earnings announcements to show that CEO gender correlates with how stock markets react to earnings surprises. There is a vast literature that has used this approach to explore a varied array of topics, including post-earnings announcement drift and limited investor attention (Bernard and Thomas 1989, MacKinlay 1997, Kothari and Warner 2007, Hirshleifer, Lim, and Teoh 2009, among others). To my knowledge, this is the first paper to take advantage of market reactions around earnings announcements to investigate the role of CEO gender.

To measure the degree of market reaction to an earnings announcement for female- and male-led firms, I run the regression

$$
\begin{aligned}
\text{AR}_{i,t} = \delta_0 + \delta_{p,m} &\times 1\{\text{Surprise}_{i,t} \geq 0\} \times \text{Surprise}_{i,t} + \\
&\delta_{p,f} \times 1\{\text{Surprise}_{i,t} \geq 0\} \times \text{Surprise}_{i,t} \times 1(\text{Female-led})_{i,t} + \\
&\delta_{n,m} \times 1\{\text{Surprise}_{i,t} < 0\} \times \text{Surprise}_{i,t} + \\
&\delta_{n,f} \times 1\{\text{Surprise}_{i,t} < 0\} \times \text{Surprise}_{i,t} \times 1(\text{Female-led})_{i,t} + \\
&\delta_1 \times 1\{\text{Surprise}_{i,t} < 0\} + \delta_2 \times 1(\text{Female-led})_{i,t} + \\
&\delta_3 \times 1\{\text{Surprise}_{i,t} < 0\} \times 1(\text{Female-led})_{i,t} + u_{i,t},
\end{aligned}
\tag{14}
$$

where $i$ represents a firm, $t$ an announcement date, $\text{AR}_{i,t}$ is the abnormal return (AR) on the announcement day. I am particularly interested in the coefficient $\delta_{n,f}$ which measures differences by CEO gender in the market reaction to negative surprises.

Table 4 displays the results from estimating the specification in (14). Panel A measures the market surprise based on the FOM score, while panel B measures the market surprise using the $\frac{CS}{PastPrices}$. Column 1 of panels A and B show baseline results. Column 2 of panels A and B control for the firm's level of market capitalization and its interaction with surprises. Finally, column 3 of panels A and B further controls for firm and announcement date fixed effects.

Table 4 depicts a pattern of reaction similar to that documented in section 3: stock markets react less strongly (less negative returns) to bad news (negative market surprises) from male-led companies relative to those from their female-led peers.[17] Specifically, the estimated value for $\delta_{n,f}$ is positive and statistically significant at least at the 10% level across all specifications. The interpretation is that after negative surprises stock markets react more, in absolute value, to news from female- relative to that from male-led companies. Across specifications, abnormal returns associated with female-led companies are more than 64% more negative $((\delta_{n,f}/\delta_{n,m})$ than those associated with male-led companies. In my preferred specification (column 3 of panel A), the average negative surprise (FOM score of -0.737) yields returns of -63.9 bps and of -142.0 bps, respectively, for male- and female-led companies.

[17] Importantly, this key result remains the same even if we condition on more recent samples, from 2000 to 2022 or from 2015 to 2022. Appendix Table A16 shows the estimates for each of these samples.

Market reaction after positive surprises — that is, good news — does not appear to differ significantly by CEO gender. Indeed, even though most estimates for $\delta_{p,f}$ are positive, they are all small and not statistically different from zero. Appendix Table A15 confirms these results when observations are value-weighted instead of equally weighted as in Table 4.[18]

## 5. The sentiment gap: disagreement with the firm's narrative

In this section, I explore analysts' reaction to news about firm performance in another context: earnings conference calls. Consistent with the results in sections 3 and 4, I find that analysts disagree less with the firm's narrative after negative market surprises for male-led companies rather than for their female-led peers. This effect is solely concentrated in male analysts — the vast majority of the analyst pool. This finding sheds light on the mechanism driving the asymmetries documented in this paper: male-led companies face less pushback as they attempt to convey bad news.

### 5.1. Text measures from earnings conference calls

To explore analysts' reaction to news about firm performance in the context of earnings conference calls, I analyse the questions that they pose to firm executives. I isolate the speeches made by analysts — that is, speakers that do not belong to the firm — from those made by firm executives in these calls and build a measure of the level of semantic disagreement between analysts and executives.

Earnings conference calls are quarterly events held by public companies (not mandatory) to go over the past earnings results and give guidance about the future. These calls are composed almost universally of two sections: a presentation section where executives discuss results and guidance; and a Q&A section where analysts and investors ask questions to firm executives. Previous literature has explored this setting to identify economically relevant

[18] It is hard to ascertain how persistent these price effects are. Indeed, in one hand, coefficients are stable and economically sizable — around 80 bps for the average negative surprise — for at least 30 days (see Appendix Figure B8). In the other hand, coefficients are not statistically different from zero for most days after an announcement (see Appendix Figure B8). At lower frequencies, using annual market-to-book ratios, it is not possible to detect any differences between male- and female-led companies (see Appendix Table A17).

content — such as identifying risk factors and economically impactful technologies (Hassan, Hollander, Lent, and Tahoun 2024, Hassan, Schreger, Schwedeler, and Tahoun *Forthcoming*, Kalyani, Bloom, Carvalho, Hassan, Lerner, and Tahoun *Forthcoming*) — and to assess the informativeness of the interactions between analysts and executives (Matsumoto, Pronk, and Roelofsen 2011, Mayew and Venkatachalam 2012, Price, Doran, Peterson, and Bliss 2012, Chen, Nagar, and Schoenfeld 2018).

To build the sentiment score of analysts, I concatenate all the speeches made by analysts in the Q&A session and count the number of positive (e.g. *improve* and *upturn*) and negative words (e.g. *deterioration* and *misconduct*). To determine which words are positive and which are negative, I use the dictionary of words built by Loughran and McDonald (2011). The sentiment score is obtained by taking the difference between the number of positive ($\text{Words}_+$) and negative words ($\text{Words}_-$), and then normalizing by the total number of words (Words):

$$\text{Sentiment} = \frac{\text{Words}_+ - \text{Words}_-}{\text{Words}}. \tag{15}$$

Next, I obtain a measure of disagreement between analysts and executives which I refer to as the sentiment gap. First, I compute the sentiment score for firm executives based on their speeches in the presentation part of earnings conference calls. I then take the absolute value of the difference between executives and analysts' sentiment scores:

$$\text{Sentiment Gap} = |\text{Analysts' Sentiment} - \text{Executives' Sentiment}|. \tag{16}$$

The last step is to compute the text measures (15) and (16) based on the gender of the analysts asking questions. To that effect, I first attribute gender to speakers in these earnings conference calls by using their first name — Appendix C.2 goes over the details. Then I concatenate the speeches for all male (female) analysts and calculate their sentiment score and their sentiment gap.

I merge these text measures, which are available from 2002 to 2022, with data on market surprise on announcement day. I do this by matching transcripts to announcement dates based on a 6-day window around each announcement day (3 days before or after). If more than one transcript is matched to the same date, I keep the one closest (in absolute value) to the announcement date. In some cases, companies hold two conference calls on the same day: one for the quarterly results, one for annual results. In this case, I keep the call regarding quarterly results.

### 5.2. Interpretation: analysts' disagreement with the firm's narrative

I argue that a positive sentiment gap indicates that analysts express disagreement with the firm's narrative about firm performance, as conveyed in executives' speeches during the presentation part of earnings conference calls.

To interpret this sentiment gap, I start by analyzing its sub-components. Appendix Table A18 shows that the average sentiment gap is positive (1.335 p.p., $p < 0.001$), driven by a negative average sentiment score for analysts (-0.315 p.p., $p < 0.001$) and a positive average sentiment score for executives (0.943 p.p., $p < 0.001$). This suggests that the sentiment gap primarily reflects the degree to which analysts' sentiment is more negative than that of executives.

To illustrate what the sentiment gap captures, consider the following examples. Words in blue represent negative sentiment, and words in green represent positive sentiment. First, imagine a student presenting at a seminar and stating:

> "*Online learning provides a convenient way for students to gain knowledge and build valuable skills, offering opportunities to those who may not have access to traditional in-person classes.*"

The sentiment score of this statement is $(3 - 0)/28 = 11\%$. Now consider a professor's response to the student:

> "*But don't limitations in engagement and challenges with self-discipline often make online learning difficult for many students?*"

The sentiment score for this response is $(0 - 3)/17 = -18\%$, creating sentiment gap of 29%.

Next, consider an example similar to an earnings conference call, where a CEO states:

> "*Despite disappointing earnings, key fundamentals remain strong, and we are optimistic that our strategic investments will drive improved profitability in the coming quarters.*"

The CEO's statement has a sentiment score of $(4-1)/23 = 13\%$. Following this, a financial analyst responds:

> "*But given the weak performance in key segments, coupled with ongoing cost pressures, I am concerned that these results signal more than a temporary setback. Can you clarify how you intend to prevent further deterioration?*"

The sentiment score of the analyst's response is $(0-4)/35 = -11\%$, resulting in a sentiment gap of 24%. In both cases — the general context with the professor and the student, and the earnings call with the analyst and CEO — the sentiment gap captures disagreement, reflecting skepticism by the second speaker toward the first speaker's statement.

In this paper's context, which examines the full presentations of executives and all questions asked by analysts, analysts' sentiment scores can be more negative than that conveyed by firm executives in their presentations, i.e. the firm's narrative, for two reasons. First, analysts may ask about negative topics more often than firm executives bring them up in their speeches. Consider the following excerpt from an actual question in an earnings conference call:[19]

> "*This week, unions for aircraft controllers, pilots and flight attendants sounded the alarm over the shutdown's effects on the U.S. aviation system and raised questions about safety, that the entire system could, at some point break.*"

In the example, if analysts ask questions about this negative topic — the shutdown — more frequently than it was brought up in the firm's narrative, then the sentiment score of analysts will be more negative than that of executives.

The second case in which the sentiment score of analysts may be more negative than that of the firm's narrative is if they use more negative words to describe a certain topic relative to firm executives. For instance, consider the following question from an actual earnings conference call:[20]

[19] Transcript for the earnings conference call for American Airlines Group Inc (Q4 2018). Part of the question asked by Andrew Tangel.

[20] Transcript for the earnings conference call for FTI Consulting (Q2 2003). Part of the question asked by Jason Malmont.

"*I am just trying to reconcile your [previous] statement (...). I just saw a report today (...) showing defaults down 10%, from their distress side we are seeing a lack of flow in terms of restructurings, a lot of stuff coming out of bankruptcy, how does that fit with your expectation that things wouldn't slow down on that side?*"

In this example, an analyst asks a question about a statement previously made by a firm executive and uses negative words to further characterize it. If this happens frequently in a call — that is, if analysts frequently use more negative language to address a certain topic compared to firm executives — then the sentiment of analysts will tend to be more negative than that of the firm's narrative.

The interpretation is therefore that a larger sentiment gap indicates that analysts disagree more with the firm's narrative, either because they ask about negative topics more often than firm executives bring them up, or because they use more negative words to characterize the same topics that were addressed by the firm's narrative.

### 5.3. Relationship between disagreement and market surprises

I show that, consistent with the results of sections 3 and 4, analysts express less disagreement with firm executives after negative market surprises in earnings conference calls from male-led companies rather than in those from their female-led peers.

I build on a literature that has explored interactions between CEO gender and participation in earnings conference calls (Klevak, Livnat, and Suslava 2024, Milian, Smith, and Alfonso 2017, Davis, Ge, Matsumoto, and Zhang 2015). De Amicis, Falconieri, and Tastan (2021) find that analysts participating in earnings conference calls with female senior managers (CEOs and CFOs) exhibit less positive sentiment than those in earnings conference calls with male senior managers. Brown, Francis, Hu, Shohfi, Zhang, and Xin (2023) find that female executives are more frequently interrupted — by both analysts and other executives — than their male peers. Comprix, Lopatta, and Tideman (2022) find that analysts display more aggressiveness in their questioning of female rather than male executives — as captured by more direct questions, a greater number of follow-up questions, larger prefaces before questions, and more questions in negative form.

I build on these papers by exploring the state-dependency in analysts' disagreement with firm executives: I show that the sentiment gap between analysts and executives is systematically correlated with both CEO gender and whether the earnings conference call is associated with positive or negative market surprises.

To explore state dependency in the sentiment gap, I run the following regression:

$$\begin{aligned}
\text{Sentiment Gap}_{i,t} = a_0 + a_1 \times 1(\text{Female-led})_{i,t} + a_{n,m} \times 1(\text{Surprise}_{i,t} < 0) + \\
a_{n,f} \times 1(\text{Female-led})_{i,t} \times 1(\text{Surprise}_{i,t} < 0) + \\
a_z \times Z_{i,t} + u_{i,t},
\end{aligned} \tag{17}$$

where $\text{Surprise}_{i,t}$ is a measure of market surprise on announcement day — I use the FOM score described in section 2 —, and $Z_{i,t}$ includes announcement date and firm fixed effects.

I am interested in the coefficient $a_{n,f}$, which tests whether the sentiment gap is different between male- and female-led companies when there is a negative market surprise. Given the finding in sections 3 and 4 that male CEOs — but not female ones — are given the benefit of the doubt after bad news, I would expect coefficient $a_{n,f}$ to be positive. The interpretation being that the benefit of the doubt reflects in the sentiment of analysts being more aligned with that of firm executives — i.e. it reflects in a lower sentiment gap.

Panel A in Table 5 displays the results, with column 2 including firm and announcement date fixed effects.

The key result is that the disagreement between analysts and executives — i.e. the sentiment gap — is lower in calls associated with negative surprises for male-led companies relative to their female-led peers. In particular, while the coefficient on negative market surprises is negative, the coefficient on the interaction between those negative surprises and female-led companies is positive and statistically significant — the respective estimates are -0.088 ($p < 0.001$) and 0.079 ($p = 0.006$). To get a sense of the magnitude of these effects, the average disagreement between analysts and firm executives — captured by the constant $a_0$ — is of 1.357 ($p < 0.001$).

Since I observe the full name of analysts in the transcripts of earnings conference calls, I can examine these results by the gender of the analyst asking questions. To that effect, I compute the sentiment gap separately for male and female analysts. Panel B in Table 5 displays the results.

The difference in the sentiment gap by CEO gender is driven by male analysts, with no such difference by CEO gender documented for female analysts.[21] For male analysts (columns 1 and 2 of panel B in Table 5), the coefficient on the interaction of negative market surprises and female CEOs is positive and statistically significant. The interpretation is that male analysts in calls associated with negative market surprises express less disagreement with the firm's narrative when there is a male CEO rather than a female CEO. For female analysts, the estimates are not suggestive of significant differences by CEO gender. The rest of the coefficients, for example the coefficient on negative surprises or the constant, are numerically similar across male and female analysts.

This asymmetry between effects by gender of the analyst asking questions suggests that the imbalance between male and female analysts matters for the overall reaction to news about firm performance documented in this paper. For instance, in the sample, female analysts represent, on average, less than 15% of the outside participants in an earnings conference call. This finding also speaks to Jannati, Kumar, Niessen-Ruenzi, and Wolfers (2023), who document differences by analyst gender in forecasts about firm performance.

An analysis of the sub-components of the sentiment gap — in Table 6 — is informative about the drivers of my results. After negative surprises, the larger sentiment gap between male analysts and executives of female-led companies (relative to their male-led peers) reflects both (i) a more negative sentiment of male analysts for these female-led firms (relative to their male-led peers), and (ii) a more positive sentiment from the executives of these female-led firms (relative to their male-led peers). The more negative sentiment of male analysts in calls with female CEOs compared to those with male CEOs is consistent with the benefit of the doubt documented in previous sections. In turn, the more positive sentiment of the firm's narrative in female-led companies relative to their male-led peers could be explained by executives anticipating the more negative sentiment of male analysts — the majority of participants in the Q&A session —, or by the lower persistence of yearly earnings growth in

[21]Note that this result remains the same even if we condition on the more recent sample of 2015 to 2022, as Appendix Table A19 shows.

these companies (relative to their male-led peers) — documented and discussed in Appendix D.[22,23]

These findings shed light on the potential mechanism for the results in this paper: there is a reduced level of skepticism, or an increased level of credulity, from male market participants toward male CEOs. In particular, my findings suggest that male CEOs face less pushback as they attempt to convey bad news. This allows them to set the narrative about the bad news more easily. This reduced skepticism is reflected in the biased pattern observed in analysts' forecasts, stock prices, and disagreement.

Note that one alternative explanation for the results in this paper is that male-led companies are better at making excuses for their bad news. In this case, we should expect a more positive sentiment from executives in male-led companies after bad news — i.e., negative earnings surprises — compared to those in female-led companies. The results in Table 6 do not support this hypothesis: the sentiment score of executives in male-led companies after negative surprises is actually more negative than that of executives in their female-led peers.

## 6. Conclusion

I have shown that there is an asymmetry in how market participants react to news about firm performance by CEO gender: market participants react less to bad news from male-led companies relative to their female-led peers, but react similarly to good news from these two types of firms. This effect is driven by an under-reaction to bad news about male-led companies relative to the rational expectations benchmark. In words: men seem to receive the benefit of the doubt from market participants after bad news. Evidence from the language used in earnings conference calls suggests that male-led companies face less pushback as they attempt to convey bad news. I argue that these patterns likely reflect frictions a reduced level of skepticism towards male CEOs.

The interpretation of the results in this paper rests on the assumption that the perfor-

[22]To see why, note that, after negative surprises, the lower persistence of earnings growth for female-led companies compared to their male-led peers suggests a more positive future outlook for these female-led companies relative to that of their male-led peers.

[23]The more positive tone of the firm's narrative in female-led companies relative to their male-led peers is in line with previous literature (e.g., De Amicis, Falconieri, and Tastan 2021).

mance of public companies is associated with the talent of their CEO. Such association is due to her holding the highest leadership position in the company, as well as to her higher visibility to the public relative to other firm executives. For instance, between 2002 and 2022, 73% of quarterly earnings conference calls of US public companies had the CEO accounting for more than 50% of all words spoken by firm executives during the Q&A session. Moreover, associating firm performance with the talent of the CEO is in line with the literature on the estimation of managerial ability (e.g., Bertrand and Schoar 2003, Demerjian, Lev, and McVay 2012, among others) and on the rise of CEO compensation (e.g. Malmendier and Tate 2009, Gabaix and Landier 2008, among others).

Using the perception of firm performance as a proxy for the perception of CEO talent, the results in this paper indicate that there is a gender-based difference in how CEO talent is perceived. My findings suggest that there might be a mechanism operating through market beliefs that further limits the advancement of women to the top of the executive ladder. In this sense, this paper contributes to the vast literature addressing the reasons for gender-based differences in labor outcomes (Goldin 2014, Buser, Nierdele, and Oosterbeek 2014, Bertrand, Goldin, and Katz 2010, Blau and Kahn 2006, among others), especially to the literature on the gender gap in executive pay (Bertrand and Hallock 2001, Gayle, Golan, and Miller 2012, among others). Future research is needed to further understand this potential belief-based mechanism to the under-representation of women in CEO positions.

# Tables

Table 1: Predictive regressions by CEO gender

**Panel A: regressions by sub-sample**

| Sub-sample: | Male-led Companies | | Female-led Companies | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Dependent Variable: | Forecast Error$_{a,i,t+1}$ | | | |
| $b_1$: Revision$_{a,i,t}$ | 0.753** | 0.515* | 0.069 | -0.284 |
| | (0.298) | (0.285) | (0.150) | (0.203) |
| Observations | 488,115 | 486,014 | 16,700 | 15,957 |
| R-squared | 0.002 | 0.134 | 0.000 | 0.271 |
| Controls | No | Yes | No | Yes |
| Forecast Period, Analyst FEs | No | Yes | No | Yes |
| Firm, Broker FEs | No | Yes | No | Yes |
| Std Errors (clustered) | Industry&Year | Industry&Year | Industry&Year | Industry&Year |

**Panel B: regressions in stacked specifications**

| | (1) | (2) | (3) |
|---|---|---|---|
| Dependent Variable: | Forecast Error$_{a,i,t+1}$ | | |
| $\beta_m$: Revision$_{a,i,t}$ | 0.753** | 0.692* | 0.515* |
| | (0.298) | (0.343) | (0.283) |
| $\beta_f$: Revision$_{a,i,t}$ x 1(Female-led)$_{i,t}$ | -0.684** | -0.684** | -0.596** |
| | (0.266) | (0.263) | (0.221) |
| Observations | 504,815 | 504,815 | 502,710 |
| R-squared | 0.002 | 0.003 | 0.133 |
| Controls | No | Yes | Yes |
| Forecast Period, Analyst FEs | No | No | Yes |
| Firm, Broker FEs | No | No | Yes |
| Std Errors (clustered) | Industry&Year | Industry&Year | Industry&Year |
| $\beta_m + \beta_f$ | 0.069 | 0.008 | -0.080 |
| $H_0$: $\beta_m + \beta_f = 0$ (p-value) | 0.646 | 0.973 | 0.730 |

Notes: Panel A presents the results from a regression of forecast errors, Forecast Error$_{a,i,t+1}$, on forecast revisions, Revision$_{a,i,t}$, for two sub-samples: companies with only male CEO/co-CEOs, and companies with at least one female CEO/co-CEO. The particular specification is: Forecast Error$_{a,i,t+1} = b_0 + b_1$Revision$_{a,i,t} + u_{a,i,t+1}$, where $a$ represents an analyst, $i$, a firm-CEO pair, and $t$, a quarter. Columns 2 and 4 control for firm-level variables (Forecast$_{a,i,t-1}$; market capitalization; an indicator of whether capitalization is lower than $2 billion, and its interaction with revisions; an indicator of whether capitalization is higher than $10 billion, and its interaction with revisions), and a series of fixed effects (forecast period, analyst, firm, and broker fixed effects). Panel B presents results for a stacked specification: Forecast Error$_{a,i,t+1} = \beta_0 + \beta_1 \times$ 1(Female-led)$_{i,t} + \beta_m \times$ Revision$_{a,i,t} + \beta_f \times$ Revision$_{a,i,t} \times$ 1(Female-led)$_{i,t} + u_{a,i,t+1}$. Columns 2 and 3 control for firm-level variables. Column 3 further adds a series of fixed effects. Market capitalization is measured two quarters before the forecast period (i.e., at t-1). Standard errors are clustered at the 4-digit SIC code and year levels. If a firm has more than one CEO in a given period, its observations are weighted by the inverse of the total number of co-CEOs at that period. In all regressions, forecast errors and revisions are normalized by the lagged end-of-quarter price of a firm's stock, $P_{i,t-1}$. Forecast revisions are winsorized at the 1% and 99% levels.

## Table 2: Predictive regressions by sign of news and CEO gender

**Panel A: predicting forecast errors after negative forecast revisions**

| Sub-sample: | Bad News *($Revision_{a,i,t} < 0$)* | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Dependent Variable: | Forecast $\text{Error}_{a,i,t+1}$ | | |
| $\beta_m$: $\text{Revision}_{a,i,t}$ | 1.072** | 0.932* | 0.812* |
| | (0.460) | (0.506) | (0.463) |
| $\beta_f$: $\text{Revision}_{a,i,t}$ x 1(Female-led)$_{i,t}$ | -1.051** | -1.064** | -0.838*** |
| | (0.409) | (0.409) | (0.272) |
| Observations | 294,916 | 294,916 | 292,483 |
| R-squared | 0.004 | 0.004 | 0.278 |
| Controls | No | Yes | Yes |
| Forecast Period, Analyst FEs | No | No | Yes |
| Firm, Broker FEs | No | No | Yes |
| Std Errors (clustered) | Industry&Year | Industry&Year | Industry&Year |
| $\beta_m + \beta_f$ | 0.021 | -0.132 | -0.026 |
| $H_0$: $\beta_m + \beta_f = 0$ (p-value) | 0.931 | 0.711 | 0.940 |

**Panel B: predicting forecast errors after positive forecast revisions**

| Sub-sample: | Good News *($Revision_{a,i,t} \geq 0$)* | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Dependent Variable: | Forecast $\text{Error}_{a,i,t+1}$ | | |
| $\beta_m$: $\text{Revision}_{a,i,t}$ | -0.020 | 0.167 | 0.302 |
| | (0.188) | (0.165) | (0.264) |
| $\beta_f$: $\text{Revision}_{a,i,t}$ x 1(Female-led)$_{i,t}$ | 0.052 | 0.088 | 0.045 |
| | (0.208) | (0.221) | (0.235) |
| Observations | 209,899 | 209,899 | 207,362 |
| R-squared | 0.000 | 0.000 | 0.159 |
| Controls | No | Yes | Yes |
| Forecast Period, Analyst FEs | No | No | Yes |
| Firm, Broker FEs | No | No | Yes |
| Std Errors (clustered) | Industry&Year | Industry&Year | Industry&Year |
| $\beta_m + \beta_f$ | 0.032 | 0.255 | 0.348 |
| $H_0$: $\beta_m + \beta_f = 0$ (p-value) | 0.852 | 0.233 | 0.370 |

Notes: Panels A and B present the results from a regression of future forecast errors, Forecast $\text{Error}_{a,i,t+1}$, on forecast revisions, $\text{Revision}_{a,i,t}$, an indicator of whether the firm has a female CEO/co-CEO, 1(Female-led)$_{i,t}$, and its interaction with forecast revisions. The particular specification is: Forecast $\text{Error}_{a,i,t+1} = \beta_0 + \beta_1 \times 1(\text{Female-led})_{i,t} + \beta_m \times \text{Revision}_{a,i,t} + \beta_f \times \text{Revision}_{a,i,t} \times 1(\text{Female-led})_{i,t} + u_{a,i,t+1}$, where $a$ represents an analyst, $i$, a firm-CEO pair, and $t$, a quarter. Panel A shows results for the sub-sample of negative forecast revisions, while panel B displays estimates for the sub-sample of positive forecast revisions. In both panels, columns 2 and 3 control for firm-level variables: Forecast(t-2); market capitalization; an indicator of whether capitalization is lower than $2 billion, and its interaction with forecast revisions; an indicator of whether capitalization is higher than $10 billion, and its interaction with forecast revisions. In both panels, column 3 further adds a series of fixed effects: forecast period, analyst, firm, and broker fixed effects. Market capitalization is measured two quarters before the forecast period quarter (i.e., at t-1). Standard errors are clustered at the 4-digit SIC code and year levels. If a firm has more than one CEO in a given period, its observations are weighted by the inverse of the total number of co-CEOs at that period. In all regressions, forecast errors and forecast revisions are normalized by the lagged end-of-quarter price of a firm's stock, $P_{i,t-1}$. Forecast revisions are winsorized at the 1% and 99% levels.

Table 3: Predictive regressions: fully stacked specification

| | (1) | (2) | (3) |
|---|---|---|---|
| | Forecast Error$_{a,i,t+1}$ | | |
| $B_{m,b}$: Revision$_{a,i,t}$ x 1(Revision$_{a,i,t}$ < 0) | 1.072** | 0.932* | 0.783* |
| | (0.460) | (0.506) | (0.428) |
| $B_{f,b}$: Revision$_{a,i,t}$ x 1(Female-led)$_{i,t}$ x 1(Revision$_{a,i,t}$ < 0) | -1.051** | -1.064** | -0.947*** |
| | (0.409) | (0.410) | (0.335) |
| $B_{m,g}$: Revision$_{a,i,t}$ x 1(Revision$_{a,i,t}$ $\geq$ 0) | -0.020 | 0.167 | 0.054 |
| | (0.188) | (0.165) | (0.090) |
| $B_{f,g}$: Revision$_{a,i,t}$ x 1(Female-led)$_{i,t}$ x 1(Revision$_{a,i,t}$ $\geq$ 0) | 0.052 | 0.089 | -0.039 |
| | (0.208) | (0.237) | (0.248) |
| Observations | 504,815 | 504,815 | 502,710 |
| R-squared | 0.003 | 0.003 | 0.134 |
| Controls | No | Yes | Yes |
| Forecast Period, Analyst FEs | No | No | Yes |
| Firm, Broker FEs | No | No | Yes |
| Std Errors (clustered) | Industry&Year | Industry&Year | Industry&Year |
| $B_{m,g} - B_{m,b}$ | -1.092 | -0.765 | -0.728 |
| $H_0: \ B_{m,g} - B_{m,b} = 0$ (p-value) | 0.080 | 0.200 | 0.078 |
| $B_{m,g} + B_{f,g} - (B_{m,b} + B_{f,b})$ | 0.011 | 0.388 | 0.180 |
| $H_0: \ B_{m,g} + B_{f,g} - (B_{m,b} + B_{f,b}) = 0$ (p-value) | 0.977 | 0.383 | 0.647 |

Notes: This table contains regressions with a fully stacked specification for panels A and B of Table 2. In particular, this table shows the results of a regression of forecast errors, Forecast Error(t+1), on: (i) a constant, (ii) an indicator for whether the company has a female CEO/co-CEO, (iii) an indicator for whether forecast revisions are negative, (iv) interactions between these variables and forecast revisions. The particular specification is: Forecast $\text{Error}_{a,i,t+1} = B_0 + B_1 \times 1(\text{Female-led})_{i,t} + B_2 \times 1(\text{Revision}_{a,i,t} < 0) + B_3 \times 1(\text{Revision}_{a,i,t} < 0) \times 1(\text{Female-led})_{i,t} + B_{m,g}\text{Revision}_{a,i,t} \times 1(\text{Revision}_{a,i,t} \geq 0) + B_{f,g} \times \text{Revision}_{a,i,t} \times 1(\text{Revision}_{a,i,t} \geq 0) \times 1(\text{Female-led})_{i,t} + B_{m,b}\text{Revision}_{a,i,t} \times 1(\text{Revision}_{a,i,t} < 0) + B_{f,b} \times \text{Revision}_{a,i,t} \times 1(\text{Revision}_{a,i,t} < 0) \times 1(\text{Female-led})_{i,t} + u_{a,i,t+1}$, where $i$ is a firm-CEO pair, $a$ an analyst, and $t$ a forecast period. Columns 2 and 3 control for market capitalization and past level of forecasts, Forecast(t-2). Column 3 further controls for a series of fixed effects: analyst, forecast period, firm, and broker fixed effects. Standard errors are clustered at the 4-digit SIC code and year levels. If a firm has more than one CEO in a given period, its observations are weighted by the inverse of the total number of co-CEOs at that period. In all regressions, forecast errors and forecast revisions are normalized by the lagged end-of-quarter price of a firm's stock, $P_{i,t-1}$. Forecast revisions are winsorized at the 1% and 99% levels.

Table 4: Stock market reaction to earnings announcements

**Panel A: using the FOM score to measure surprises**

| | (1) | (2) | (3) |
|---|---|---|---|
| Dependent Variable: | | Abnormal Return$_{i,t}$ | |
| $\delta_{n,m}$: Surprise$_{i,t}$ x 1(Surprise$_{i,t}$ < 0) | 1.737*** | 1.593*** | 1.602*** |
| | (0.127) | (0.182) | (0.201) |
| $\delta_{n,f}$: Surprise$_{i,t}$ x 1(Surprise$_{i,t}$ < 0) x 1(Female-led)$_{i,t}$ | 1.111** | 1.101** | 1.284** |
| | (0.458) | (0.464) | (0.494) |
| $\delta_{p,m}$: Surprise$_{i,t}$ x 1(Surprise$_{i,t}$ ≥ 0) | 1.497*** | 1.125*** | 1.108*** |
| | (0.071) | (0.140) | (0.151) |
| $\delta_{p,f}$: Surprise$_{i,t}$ x 1(Surprise$_{i,t}$ ≥ 0) x 1(Female-led)$_{i,t}$ | 0.066 | 0.063 | 0.002 |
| | (0.228) | (0.228) | (0.211) |
| Observations | 232,775 | 229,562 | 228,376 |
| R-squared | 0.039 | 0.040 | 0.143 |
| Firm-level Controls | No | Yes | Yes |
| Firm, Announcement Date FEs | No | No | Yes |
| Cluster Level | Industry&Year | Industry&Year | Industry&Year |
| $\delta_{p,m} - \delta_{n,m}$ | -0.240 | -0.469 | -0.494 |
| $H_0: \delta_{p,m} - \delta_{n,m} = 0$ (p-value) | 0.058 | 0.019 | 0.037 |
| $\delta_{p,m} + \delta_{p,f} - (\delta_{n,m} + \delta_{n,f})$ | -1.285 | -1.507 | -1.776 |
| $H_0: \delta_{p,m} + \delta_{p,f} - (\delta_{n,m} + \delta_{n,f}) = 0$ (p-value) | 0.009 | 0.004 | 0.001 |

**Panel B: using the $CS/PastPrices$ to measure surprises**

| | (1) | (2) | (3) |
|---|---|---|---|
| Dependent Variable: | | Abnormal Return$_{i,t}$ | |
| $\delta_{n,m}$: Surprise$_{i,t}$ x 1(Surprise$_{i,t}$ < 0) | 0.087*** | 0.076** | 0.027 |
| | (0.024) | (0.031) | (0.036) |
| $\delta_{n,f}$: Surprise$_{i,t}$ x 1(Surprise$_{i,t}$ < 0) x 1(Female-led)$_{i,t}$ | 0.148** | 0.142* | 0.177* |
| | (0.071) | (0.076) | (0.091) |
| $\delta_{p,m}$: Surprise$_{i,t}$ x 1(Surprise$_{i,t}$ ≥ 0) | 0.249*** | 0.122 | 0.243** |
| | (0.046) | (0.085) | (0.105) |
| $\delta_{p,f}$: Surprise$_{i,t}$ x 1(Surprise$_{i,t}$ ≥ 0) x 1(Female-led)$_{i,t}$ | 0.007 | 0.005 | -0.089 |
| | (0.111) | (0.112) | (0.097) |
| Observations | 232,755 | 229,559 | 228,373 |
| R-squared | 0.032 | 0.033 | 0.136 |
| Firm-level Controls | No | Yes | Yes |
| Firm, Announcement Date FEs | No | No | Yes |
| Cluster Level | Industry&Year | Industry&Year | Industry&Year |
| $\delta_{p,m} - \delta_{n,m}$ | 0.162 | 0.046 | 0.216 |
| $H_0: \delta_{p,m} - \delta_{n,m} = 0$ (p-value) | 0.002 | 0.618 | 0.065 |
| $\delta_{p,m} + \delta_{p,f} - (\delta_{n,m} + \delta_{n,f})$ | 0.021 | -0.091 | -0.050 |
| $H_0: \delta_{p,m} + \delta_{p,f} - (\delta_{n,m} + \delta_{n,f}) = 0$ (p-value) | 0.861 | 0.490 | 0.658 |

Notes: The dependent variable in regressions is the abnormal return (AR) over the S&P 500 index on the earnings announcement day. In both panels, the regression specification is: $\text{AR}_{i,t} = \delta_0 + \delta_{m,p} \times 1\{\text{Surprise}_{i,t} \geq 0\} \times \text{Surprise}_{i,t} + \delta_{f,p} \times 1\{\text{Surprise}_{i,t} \geq 0\} \times \text{Surprise}_{i,t} \times 1(\text{Female-led})_{i,t} + \delta_{m,n} \times 1\{\text{Surprise}_{i,t} < 0\} \times \text{Surprise}_{i,t} + \delta_{f,n} \times 1\{\text{Surprise}_{i,t} < 0\} \times \text{Surprise}_{i,t} \times 1(\text{Female-led})_{i,t} + \delta_1 \times 1\{\text{Surprise}_{i,t} < 0\} + \delta_2 \times 1(\text{Female-led})_{i,t} + \delta_3 \times 1\{\text{Surprise}_{i,t} < 0\} \times 1(\text{Female-led})_{i,t} + u_{i,t}$, where $i$ is a firm-CEO pair and $t$, an announcement date. The measure of surprise in panel A is the FOM score, following the methodology in Chiang, Dai, Fan, Hong, and Tu (2019), while the measure of surprise in panel B is the average surprise across analysts (consensus surprise, CS) normalized by the firm's stock price at $t - 20$ days. In both panels, column 2 controls for: the firm's market capitalization in $t - 20$ days, indicators for whether market capitalization is lower than \$2 billion and whether it is higher than \$10 billion, and their interactions with surprises. Column 3 further controls for announcement date and firm fixed effects. Standard errors are clustered at the 4-digit SIC code and year of announcement levels. If a firm has more than one CEO in a given period, its observations are weighted by the inverse of the total number of co-CEOs at that period.

Table 5: Sentiment gap in earnings conference calls by CEO gender

**Panel A: Disagreement between analysts and firm executives**

| | (1) | (2) |
|---|---|---|
| Dependent Variable: | Sentiment $\text{Gap}_{i,t}$ | |
| $a_{n,f}$: 1($\text{Surprise}_{i,t} < 0$) x 1(Female-led)$_{i,t}$ | 0.115** | 0.079*** |
| | (0.048) | (0.025) |
| $a_{n,m}$: 1($\text{Surprise}_{i,t} < 0$) | -0.135*** | -0.088*** |
| | (0.013) | (0.010) |
| $a_1$: 1(Female-led)$_{i,t}$ | 0.054 | 0.014 |
| | (0.043) | (0.037) |
| $a_0$: Constant | 1.369*** | 1.357*** |
| | (0.028) | (0.003) |
| Observations | 77,243 | 76,549 |
| R-squared | 0.006 | 0.323 |
| Firm, Announcement Date FEs | No | Yes |
| Std Errors (clustered) | Industry&Year | Industry&Year |

**Panel B: Disagreement by analyst gender**

| Sub-sample: | Male Analysts | | Female Analysts | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Dependent Variable: | Sentiment $\text{Gap}_{i,t}$ | | | |
| $a_{n,f}$: 1($\text{Surprise}_{i,t} < 0$) x 1(Female-led)$_{i,t}$ | 0.132** | 0.098*** | 0.064 | -0.018 |
| | (0.053) | (0.033) | (0.069) | (0.074) |
| $a_{n,m}$: 1($\text{Surprise}_{i,t} < 0$) | -0.135*** | -0.087*** | -0.090*** | -0.061*** |
| | (0.014) | (0.011) | (0.016) | (0.018) |
| $a_1$: 1(Female-led)$_{i,t}$ | 0.062 | 0.006 | -0.072 | -0.001 |
| | (0.045) | (0.041) | (0.069) | (0.072) |
| $a_0$: Constant | 1.371*** | 1.360*** | 1.728*** | 1.720*** |
| | (0.028) | (0.003) | (0.024) | (0.004) |
| Observations | 77,057 | 76,357 | 36,828 | 35,901 |
| R-squared | 0.006 | 0.309 | 0.001 | 0.219 |
| Firm, Announcement Date FEs | No | Yes | No | Yes |
| Std Errors (clustered) | Industry&Year | Industry&Year | Industry&Year | Industry&Year |

Notes: Columns 1 and 2 of panel A present the results from regressions of a measure of disagreement — the distance between analysts' and executives' sentiment scores (the sentiment gap) — for a given firm and earnings call on indicators for negative surprises and for whether the firm has a female CEO/co-CEO, and their interaction. The specification is: Sentiment $\text{Gap}_{i,t} = a_0 + a_1 \times 1(\text{Female-led})_{i,t} + a_{n,m} \times 1(\text{Surprise}_{i,t} < 0) + a_{n,f} \times 1(\text{Female-led})_{i,t} \times 1(\text{Surprise}_{i,t} < 0) + u_{i,t}$, where $i$ is a firm-CEO pair and $t$ is an announcement date. Column 2 includes firm and announcement date fixed effects. Panel B presents the results for the same specifications in columns 1 and 2 of panel A, but broken down by the gender of analysts asking questions. Columns 1 and 2 consider only text from male analysts when computing the sentiment gap, while columns 3 and 4 consider only text from female analysts. Standard errors are clustered at the 4-digit SIC code and year levels. If a firm has more than one CEO in a given period, its observations are weighted by the inverse of the total number of co-CEOs at that period.

Table 6: Sentiment gap in earnings conference calls by sub-component

| Sub-sample: | Male Analysts | Female Analysts | Executives |
|---|---|---|---|
| | (1) | (2) | (3) |
| Dependent Variable: | Sentiment$_{i,t}$ | | |
| 1(Surprise$_{i,t}$ < 0) x 1(Female-led)$_{i,t}$ | -0.038* | 0.165* | 0.085*** |
| | (0.022) | (0.084) | (0.031) |
| 1(Surprise$_{i,t}$ < 0) | -0.139*** | -0.145*** | -0.248*** |
| | (0.008) | (0.022) | (0.017) |
| 1(Female-led)$_{i,t}$ | 0.007 | -0.029 | 0.008 |
| | (0.018) | (0.063) | (0.036) |
| Constant | -0.268*** | -0.367*** | 1.011*** |
| | (0.002) | (0.005) | (0.004) |
| Observations | 76,357 | 35,901 | 76,549 |
| R-squared | 0.266 | 0.220 | 0.510 |
| Firm, Announcement Date FEs | Yes | Yes | Yes |
| Std Errors (clustered) | Industry&Year | Industry&Year | Industry&Year |

Notes: The columns in this table present the results from regressions of sentiment on indicators for negative surprises and for whether the firm has a female CEO/co-CEO, and their interaction. The specification is: $\text{Sentiment}_{i,t} = a_0 + a_1 \times 1(\text{Female-led})_{i,t} + a_{n,m} \times 1(\text{Surprise}_{i,t} < 0) + a_{n,f} \times 1(\text{Female-led})_{i,t} \times 1(\text{Surprise}_{i,t} < 0) + u_{i,t}$, where $i$ is a firm-CEO pair and $t$ is an announcement date. All columns include firm and announcement date fixed effects. The dependent variable in column 1 is the sentiment score based on text from male analysts in the Q&A session of earnings conference calls, while in column 2 it is the sentiment based on the text from female analysts. In column 3, the dependent variable is the sentiment based on the speeches of executives in the presentation part of earnings conference call. Standard errors are clustered at the 4-digit SIC code and year levels. If a firm has more than one CEO in a given period, its observations are weighted by the inverse of the total number of co-CEOs at that period.

# Figures

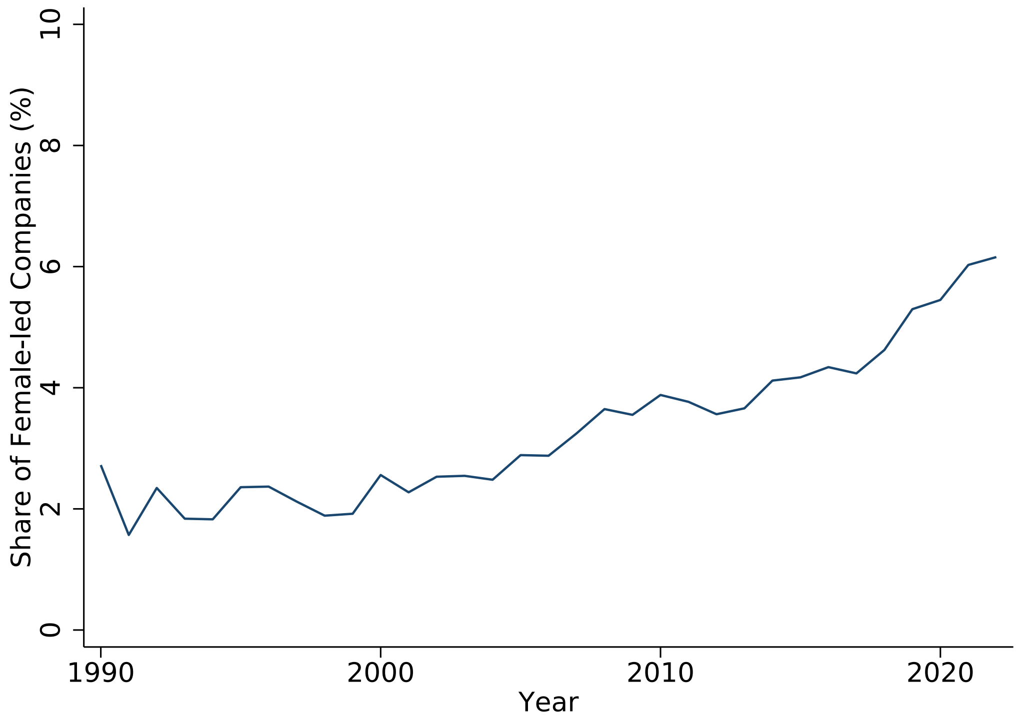



Figure 1: Share of female-led public companies in the US over time

Notes: This figure shows the average share of female-led companies over time. The sample is the universe of US public companies in IBES from January 1990 to October 2022 that satisfy the following conditions: (i) stock prices exceed $1, (ii) have non-missing information for forecast errors and revisions, (iii) have information on CEO gender following the procedure described in section 2.

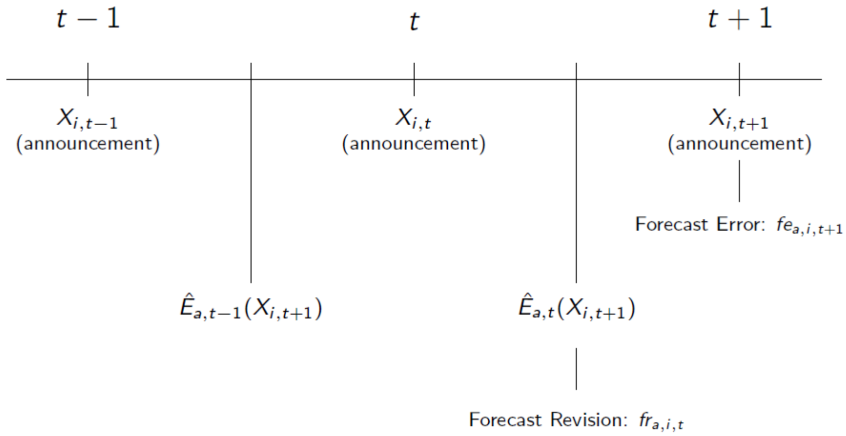
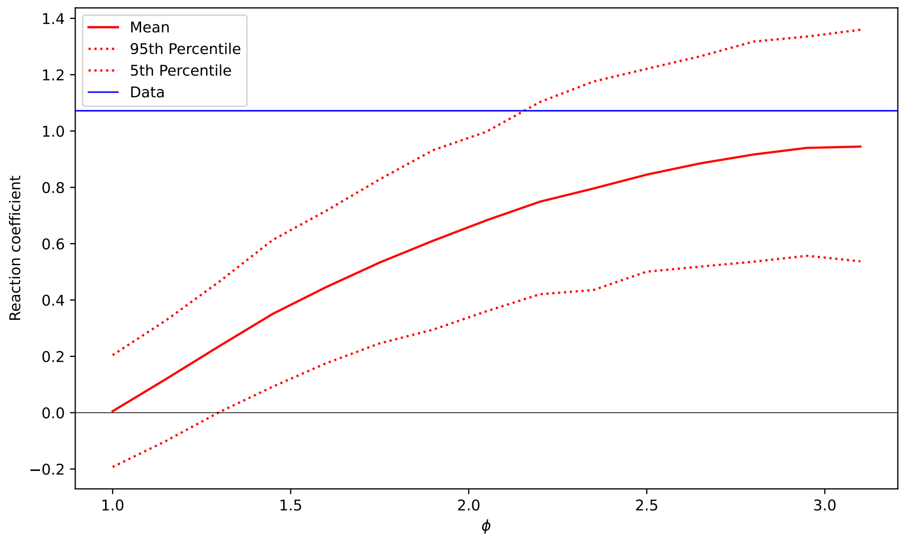
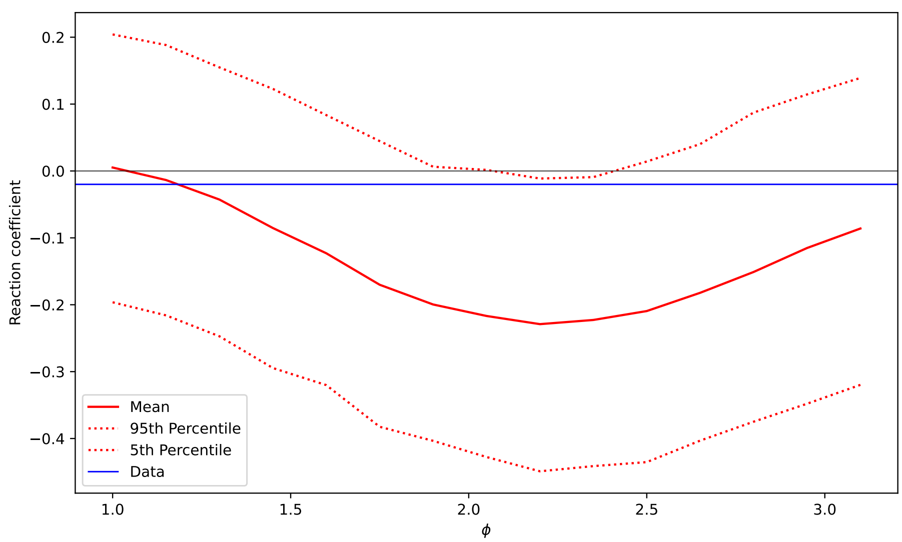



Figure 2: Timeline of earnings announcements, forecast errors and revisions




(a) Negative Forecast Revisions




(b) Positive Forecast Revisions

Figure 3: Simulating the model of section 3.3: under-estimation of the precision of signals in the bad state of the world

Notes: Panels (a) and (b) show the results of 1,000 simulations of the model in section 3.3 with 1,000 observations each. Each simulation considers a different level of under-estimation of the precision of signals in the bad state of the world: each simulation considers a different $\phi$, where $\hat{\sigma}_{\eta,\theta_{bad}} = \phi \cdot \sigma_\eta$. The simulations are based on the following parametrization: $p = 0.5$, $\mu_{bad} = 0$, $\mu_{good} = 0.144$, and $\sigma_v^2 = \sigma_\eta^2 = 0.125/2$. Panel (a) shows the mean and 90% confidence interval for the reaction coefficient of a regression of forecast errors on forecast revisions after negative forecast revisions. The blue line in panel (a) marks the point estimate for the reaction coefficient after negative forecast revisions found in the data (column 1 of panel A in Table 2). Panel (b) shows the mean and 90% confidence interval for the reaction coefficient of a regression of forecast errors on forecast revisions after positive forecast revisions. The blue line in panel (a) marks the point estimate for the reaction coefficient after positive forecast revisions found in the data (column 1 of panel B in Table 2).

## References

Albanesi, Stefania, Claudia Olivetti, and Maria José Prados (2015). "Gender and Dynamic Agency: Theory and Evidence of the Compensation of Top Executives". Federal Reserve Bank of New York Staff Reports, no. 718.

Augenblick, Ned, Eben Lazarus, and Michael Thaler (*Forthcoming*). "Overinference from Weak Signals and Underinference from Strong Signals". In: *The Quarterly Journal of Economics*.

Ball, Ray and Philip Brown (1968). "An Empirical Evaluation of Accounting Income Numbers". In: *Journal of Accounting Research* 6.2, pp. 159–178. ISSN: 00218456, 1475679X.

Bernard, Victor L. and Jacob K. Thomas (1989). "Post-Earnings-Announcement Drift: Delayed Price Response or Risk Premium". In: *Journal of Accounting Research* 27, pp. 1–36.

— (1990). "Evidence that stock prices do not fully reflect the implications of current earnings for future earnings". In: *Journal of Accounting and Economics* 13.4, pp. 305–340. ISSN: 0165-4101.

Bertrand, Marianne, Claudia Goldin, and Lawrence F. Katz (2010). "Dynamics of the Gender Gap for Young Professionals in the Financial Corporate Sectors". In: *American Economic Review: Applied Economics* 2, pp. 228–255.

Bertrand, Marianne and Kevin F. Hallock (2001). "The Gender Gap in Top Corporate Jobs". In: *Industrial and Labor Relations Review* 55.1, pp. 3–21.

Bertrand, Marianne and Antoinette Schoar (2003). "Managing With Style: The Effect Of Managers on Firm Policies". In: *The Quarterly Journal of Economics* 118 (4), pp. 1169–1208.

Blau, Francine D. and Lawrence M. Kahn (2006). "The Gender Pay Gap: Going, Going... But Not Gone". In: *The Declining Significance of Gender?* Russell Sage Foundation, pp. 37–66.

Bordalo, Pedro, Giovanni Burro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer (*Forthcoming*). "Imagining the Future: Memory, Simulation, and Beliefs". In: *Review of Economic Studies*.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer (2016). "Stereotypes". In: *The Quarterly Journal of Economics* 131.4, pp. 1753–1794.
— (2019). "Beliefs about Gender". In: *American Economic Review* 109 (3), pp. 739–773.
Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer (2019). "Diagnostic Expectations and Stock Returns". In: *The Journal of Finance* 74.6, pp. 2839–2874.
— (2022). "Belief Overreaction and Stock Market Puzzles". NBER Working Paper No. 27283.
Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer (2020). "Overreaction in Macroeconomic Expectations". In: *American Economic Review* 110.9, pp. 2748–2782.
Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2022). "Overreaction and Diagnostic Expectations in Macroeconomics". In: *Journal of Economic Perspectives* 36.3, pp. 223–244.
Bouchad, Jean-Philippe, Philipp Krüger, Augustin Landier, and David Thesmar (2019). "Sticky Expectations and the Profitability Anomaly". In: *The Journal of Finance* 74.2, pp. 639–674.
Brown, Nerissa, Bill B. Francis, Wenyao Hu, Thomas Shohfi, Tengfei Zhang, and Daqi Xin (2023). "Gender and Earnings Conference Calls". Working Paper. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3473266.
Buser, Thomas, Muriel Nierdele, and Hessel Oosterbeek (2014). "Gender, Competitiveness, and Career Choices". In: *The Quarterly Journal of Economics* 129 (3), pp. 1409–1447.
Bybee, Leland, Bryan Kelly, Asaf Manela, and Dacheng Xiu (2021). "Business News and Business Cycles". NBER Working Paper No. 29344.
Chan, Louis K. C., Narasimhan Jegadeesh, and Josef Lakonishok (1996). "Momentum Strategies". In: *The Journal of Finance* 51.5, pp. 1681–1713.
Chen, Jason V., Venky Nagar, and Jordan Schoenfeld (2018). "Manager-analyst Conversations in Earnings Conference Calls". In: *Review of Accounting Studies* 23, pp. 1315–1354.
Chiang, Chin-Han, Wei Dai, Jianqing Fan, Harrison Hong, and Jun Tu (2019). "Robust Measures of Earnings Surprises". In: *The Journal of Finance* 74.2, pp. 943–983.
Coffman, Katherine, Clio Bryant Flikkema, and Olga Shurchkov (2021). "Gender Stereotypes in Deliberation and Team Decisions". In: *Games and Economic Behavior* 129, pp. 329–349.

Coibion, Olivier and Yuriy Gorodnichenko (2015). "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts". In: *American Economic Review* 105.8, pp. 2644–2678.

Comprix, Joseph, Kerstin Lopatta, and Sebastian A. Tideman (2022). "The Role of Gender in the Aggressive Questioning of CEOs During Earnings Conference Calls". In: *The Accounting Review* 97.7, pp. 79–107.

Davis, Angela K., Weili Ge, Dawn Matsumoto, and Jenny Li Zhang (2015). "The Effect of Manager-specific Optimism on the Tone of Earnings Conference Calls". In: *Review of Accounting Studies* 20, pp. 639–673.

De Amicis, Chiara, Sonia Falconieri, and Mesut Tastan (2021). "Sentiment Analysis and Gender Differences in Earnings Conference Calls". In: *Journal of Corporate Finance* 71, pp. 3001–3025.

Demerjian, Peter, Baruch Lev, and Sarah McVay (2012). "Quantifying Managerial Ability: A New Measure and Validity Tests". In: *Management Science* 58.7, pp. 1229–1248.

Egan, Mark, Gregor Matvos, and Amit Seru (2022). "When Harry Fired Sally: The Double Standard in Punishing Misconduct". In: *Journal of Political Economy* 130.5, pp. 1184–1248.

Farmer, Leland E., Emi Nakamura, and Jón Steinsson (2024). "Learning about the Long Run". In: *Journal of Political Economy* 132.10, pp. 3334–3377.

Flynn, Joel P. and Karthik A. Sastry (2022). "The Macroeconomics of Narratives". Working Paper.

Foster, George, Chris Olsen, and Terry Shevlin (1984). "Earnings Releases, Anomalies, and the Behavior of Security Returns". In: *The Accounting Review* 59.4, pp. 574–603.

Gabaix, Xavier and Augustin Landier (2008). "Why Has CEO Pay Increased So Much?" In: *The Quarterly Journal of Economics* 123 (1), pp. 49–100.

Garcia, Diego (2013). "Sentiment during Recessions". In: *The Journal of Finance* 68.3, pp. 1267–1300.

Gayle, George-Levi, Limor Golan, and Robert A. Miller (2012). "Gender Differences in Executive Compensation and Job Mobility". In: *Journal of Labor Economics* 30.4, pp. 829–872.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy (2019). "Text as Data". In: *Journal of Economic Literature* 57 (3), pp. 535–574.

Goldin, Claudia (2014). "A Grand Gender Convergence: Its Last Chapter". In: *American Economic Review* 104.4, pp. 1091–1119.

Hassan, Tarek A., Stephan Hollander, Laurence van Lent, and Ahmed Tahoun (2019). "Firm-Level Political Risk: Measurement and Effects". In: *The Quarterly Journal of Economics* 134.4, pp. 2135–2202. ISSN: 0033-5533.

Hassan, Tarek A., Stephen Hollander, Laurence van Lent, and Ahmed Tahoun (2024). "The Global Impact of Brexit Uncertainty". In: *The Journal of Finance* 79.1, pp. 413–458.

Hassan, Tarek A., Jesse Schreger, Markus Schwedeler, and Ahmed Tahoun (*Forthcoming*). "Sources and Transmission of Country Risk". In: *The Review of Economic Studies* rdad080.

Hirshleifer, David, Sonya S. Lim, and Siew H. Teoh (2009). "Driven to Distraction: Extraneous Events and Underreaction to Earnings News". In: *The Journal of Finance* 64.5, pp. 2289–2325.

Jannati, Sima, Alok Kumar, Alexandra Niessen-Ruenzi, and Justin Wolfers (2023). "In-Group Bias in Financial Markets". Working Paper, SSRN 2884218.

Jegadeesh, Narasimhan and Sheridan Titman (1993). "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency". In: *The Journal of Finance* 48.1, pp. 65–91.

Jiang, Fuwei, Joshua Lee, Xiumin Martin, and Guofu Zhou (2019). "Manager sentiment and stock returns". In: *Journal of Financial Economics* 132, pp. 126–149.

Kalyani, Aakash, Nicholas Bloom, Marcela Carvalho, Tarek A. Hassan, Josh Lerner, and Ahmed Tahoun (*Forthcoming*). "The Diffusion of New Technologies". In: *The Quarterly Journal of Economics*.

Kaplan, Steven N. and Morten Sorensen (2021). "Are CEOs Different?" In: *The Journal of Finance* 76 (4), pp. 1607–2081.

Kelly, Bryan, Semyon Malamud, Emil Siriwardane, and Hongyu Wu (2024). "Behavioral Impulse Responses". Working Paper. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5003581.

Klevak, Julia, Joshua Livnat, and Kate Suslava (2024). "Gender Difference in CFO Communication: Evidence from Earnings Call". In: *The Accounting Review* 44, pp. 1–22.

Kothari, S. P. and Jerold B. Warner (2007). "Driven to Distraction: Extraneous Events and Underreaction to Earnings News". In: *Handbook of Empirical Corporate Finance* 1, pp. 3–36.

Kučinskas, Simas and Florian S. Peters (2022). "Measuring Under- and Overreaction in Expectation Formation". In: *The Review of Economics and Statistics*, pp. 1–45.

Landsman, Rachel (2019). "Gender Differences in Executive Departure". Working Paper.

Lee, Peggy M. and Erika Hayes James (2007). "She'-E-OS: Gender Effects and Investor Reactions to the Announcements of Top Executive Appointments". In: *Strategic Management Journal* 28.3, pp. 227–241.

Loughran, Tim and Bill McDonald (2011). "When is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks". In: *The Journal of Finance* 66.1, pp. 35–65.

MacKinlay, A. Craig (1997). "Event Studies in Economics and Finance". In: *Journal of Economic Literature* 35, pp. 13–39.

Malmendier, Ulrike and Geoffrey Tate (2009). "Superstar CEOs". In: *The Quarterly Journal of Economics* 118 (4), pp. 1593–1638.

Matsumoto, Dawn, Maarten Pronk, and Erik Roelofsen (2011). "What Makes Conference Calls Useful? The Information Content of Managers' Presentations and Analysts' Discussion Sessions". In: *The Accounting Review* 86.4, pp. 1383–1414.

Mayew, William J. and Mohan Venkatachalam (2012). "The Power of Voice: Managerial Affective States and Future Firm Performance". In: *The Journal of Finance* 67 (1), pp. 1–43.

Milian, Jonathan A., Antoinette L. Smith, and Elio Alfonso (2017). "Does an Analyst's Access to Information Vary with the Favorableness of their Language when Speaking to Management?" In: *Accounting Horizons* 31.4, pp. 13–31.

Niessen-Ruenzi, Alexandra and Stefan Ruenzi (2019). "Sex Matters: Gender Bias in the Mutual Fund Industry". In: *Management Science* 65.7, pp. 3001–3025.

Payne, Jeff L. and William B. Thomas (2003). "The Implications of Using Stock-Split Adjusted I/B/E/S Data in Empirical Research". In: *The Accounting Review* 78.4, pp. 1049–1067.

Price, S. McKay, James S. Doran, David R. Peterson, and Barbara A. Bliss (2012). "Earnings Conference Calls and Stock Returns: The Incremental Informativeness of Textual Tone". In: *Journal of Banking and Finance* 36, pp. 992–1011.

Sarsons, Heather (2024). "Interpreting Signals in the Labor Market: Evidence from Medical Referrals". Working Paper.

Sarsons, Heather, Klarita Gërxhani, Ernesto Reuben, and Arthur Schram (2021). "Gender Differences in Recognition for Group Work". In: *Journal of Political Economy* 129.1, pp. 101–147.

Tetlock, Paul C., Maytal Saar-Tsechansky, and Sofus Macskassy (2008). "More Than Words: Quantifying Language to Measure Firm's Fundamentals". In: *The Journal of Finance* 63.3, pp. 1437–1467.

Wolfers, Justin (2006). "Diagnosing Discrimination: Stock Returns and CEO Gender". In: *Journal of the European Economic Association* 4.2-3, pp. 531–541.

# A. Appendix Tables

Table A1: Summary statistics for the main sample

| Statistic: | Mean | Mean | - |
|---|---|---|---|
| | (1) | (2) | (3) |
| Sub-sample: | Male-led | Female-led | (1)-(2) |
| Market Capitalization | $8.202 bi | $8.228 bi | -0.258 |
| Price | $125.379 | $53.728 | 71.651** |
| Number of Analysts | 3.101 | 2.599 | 0.502*** |
| CEO Tenure (years) | 8.174 | 6.194 | 1.981*** |
| CEO Age (years) | 55.335 | 53.937 | 1.398*** |
| Number of Quarters | 26.968 | 20.099 | 6.069*** |
| Number of Firms | 7,892 | 540 | - |

Notes: This table shows summary statistics at the CEO level, for CEOs at male- and at female-led companies (i.e. companies with at least one female CEO/co-CEO). Column 1 shows the mean across CEOs in male-led companies, while column 2 shows the mean across CEOs in female-led companies. Column 3 shows the difference between columns 1 and 2, and whether that difference is statistically different from zero using a simple t-test. Row "Number of Firms", in column 1, shows the total number of firms that, at some point in time, were categorized as male-led companies. Row "Number of Firms", in column 2, shows the total number of firms that, at some point in time, were categorized as female-led companies. Statistics are shown for the main sample (sample of section 3): the universe of US firms at IBES whose CEOs were attributed gender, whose stock price is larger than $1, and for which forecast errors (Forecast Error(t+1)), forecast revisions (Revision(t)) and past forecasts (Forecast(t-1)) are all available.

Table A2: Frequency of revisions at the analyst-level by CEO gender

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Days until revision | | Number of revisions | |
| 1(Female-led) | -1.140** | -0.321 | -0.045** | -0.046** |
| | (0.486) | (0.304) | (0.020) | (0.020) |
| Constant | 11.710*** | 12.052*** | 1.442*** | 1.468*** |
| | (0.733) | (0.471) | (0.029) | (0.018) |
| Observations | 1,348,841 | 1,318,864 | 1,348,841 | 1,318,864 |
| R-squared | 0.000 | 0.099 | 0.000 | 0.186 |
| Controls | No | Yes | No | Yes |
| Firm, Forecast Period FEs | No | Yes | No | Yes |
| Analyst, Broker FEs | No | Yes | No | Yes |
| Std Errors (clustered) | Industry&Year | Industry&Year | Industry&Year | Industry&Year |

Notes: This table contains regressions at the analyst-level for different measures of frequency of revisions on an indicator of whether the firm has a female CEO/co-CEO. The particular specification is: $\text{Frequency}_{a,i,t} = c_0 + c_1 \times 1(\text{Female-led})_{a,i,t} + u_{a,i,t}$, where $i$ is a firm-CEO pair, $a$ an analyst, and $t$ a forecast period. In columns 1 and 2, the frequency of revisions is measured by the number of days until an analyst makes a revision after an announcement date. In columns 3 and 4, the measure of frequency is the number of revisions an analyst makes between two announcement dates. Columns 2 and 4 include firm, forecast period, analyst, and broker fixed effects. Columns 2 and 4 also control for a firm's market capitalization, CEO age and CEO tenure. If a firm has more than one CEO in a given period, its observations are weighted by the inverse of the total number of co-CEOs at that period. Standard errors are clustered at the 4-digit SIC code and year levels.

Table A3: Moments of firm performance in male- and female-led companies

**Panel A: firm performance for male- and female-led companies**

| | (1) | (2) | (3) |
|---|---|---|---|
| Measure of Performance: | QoQ Growth$_{i,t}$ | YoY Growth$_{i,t}$ | $\frac{EPS}{price}_{i,t}$ |
| $\theta_f$: 1(Female-led)$_{i,t}$ | 0.031 | -0.008 | -0.001 |
| | (0.025) | (0.032) | (0.006) |
| $\theta_m$: Constant | 0.227*** | 0.322*** | 0.052*** |
| | (0.001) | (0.001) | (0.000) |
| Observations | 129,801 | 127,262 | 164,289 |
| R-squared | 0.119 | 0.126 | 0.601 |
| Firm, Quarter-Year FEs | Yes | Yes | Yes |
| Std Errors (clustered) | Industry&Year | Industry&Year | Industry&Year |

**Panel B: volatility in firm performance for male- and female-led companies**

| | (1) | (2) | (3) |
|---|---|---|---|
| Measure of Performance: | sd(QoQ Growth)$_{i,j}$ | sd(YoY Growth)$_{i,j}$ | sd($\frac{EPS}{price}$)$_{i,j}$ |
| $\kappa_f$: 1(Female-led)$_{i,j}$ | -0.017 | 0.027 | 0.005 |
| | (0.069) | (0.073) | (0.006) |
| $\kappa_m$: Constant | 0.736*** | 0.748*** | 0.040*** |
| | (0.005) | (0.007) | (0.000) |
| Observations | 6,596 | 6,546 | 8,262 |
| R-squared | 0.605 | 0.524 | 0.610 |
| Firm FEs | Yes | Yes | Yes |
| Std Errors (clustered) | Industry | Industry | Industry |

**Panel C: persistence in firm performance for male- and female-led companies**

| | (1) | (2) | (3) |
|---|---|---|---|
| Measure of Performance: | QoQ Growth$_{i,t}$ | YoY Growth$_{i,t}$ | $\frac{EPS}{price}_{i,t}$ |
| $\varphi_f$: Performance$_{i,t-1}$ × 1(Female-led)$_{i,t}$ | -0.010 | -0.063** | -0.018 |
| | (0.017) | (0.030) | (0.033) |
| $\varphi_m$: Performance$_{i,t-1}$ | -0.184*** | 0.234*** | 0.619*** |
| | (0.009) | (0.017) | (0.035) |
| Observations | 117,410 | 113,017 | 155,747 |
| R-squared | 0.142 | 0.177 | 0.755 |
| Firm, Quarter-Year FEs | Yes | Yes | Yes |
| Std Errors (clustered) | Industry&Year | Industry&Year | Industry&Year |

Notes: Panel A shows the results of regressions of some measure of firm performance, $Y_{i,t}$ according to the specification: $Y_{i,t} = \theta_m + \theta_f 1(\text{Female-led})_{i,t} + u_{i,t}$, where $1(\text{Female-led})_{i,t}$ is an indicator for whether the company has a female CEO/co-CEO, $i$ is a firm-CEO pair, and $t$ is a quarter. All columns include quarter-year fixed effects and firm fixed effects. Panel B explores the volatility of firm performance. To that effect, I first compute the standard deviation of a certain measure of firm performance across all quarters associated with a given firm-CEO pair (firm $i$ and CEO $j$). Then I run the regression: $\text{sd}(Y_{i,j}) = \kappa_m + \kappa_f 1(\text{Female-led})_{i,j} + u_{i,j}$. All columns include firm fixed effects. Panel C presents the results from an autoregressive regression for different measures of firm performance, following the specification: $Y_{i,t} = \varphi_0 + \varphi_1 1(\text{Female-led})_{i,t} + \varphi_m Y_{i,t-1} + \varphi_f Y_{i,t-1} \times 1(\text{Female-led})_{i,t} + u_{i,t}$. All columns include firm fixed effects and quarter-year fixed effects. If a firm has more than one CEO in a given period, its observations are weighted by the inverse of the total number of co-CEOs at that period. Standard errors are clustered at the 4-digit SIC code and year levels for panels A and C. In panel B, standard errors are clustered at the 4-digit SIC code. All measures of firm performance are winsorized at the 1% and 99% levels.

Table A4: Predictive regressions: other definitions of good and bad news

| **Panel A: predicting forecast errors by sign of past forecast errors** | | | | |
|---|---|---|---|---|
| Sub-sample: | Forecast $\text{Error}_{a,i,t} < 0$ | | Forecast $\text{Error}_{a,i,t} \geq 0$ | |
| | (1) | (2) | (3) | (4) |
| Dependent variable: | Forecast $\text{Error}_{a,i,t+1}$ | | | |
| $\beta_m$: $\text{Revision}_{a,i,t}$ | 0.971** | 0.726* | 0.314*** | 0.267* |
| | (0.413) | (0.411) | (0.092) | (0.144) |
| $\beta_f$: $\text{Revision}_{a,i,t}$ x $1(\text{Female-led})_{i,t}$ | -0.970** | -0.644** | -0.104 | -0.140 |
| | (0.407) | (0.290) | (0.235) | (0.204) |
| Observations | 134,103 | 131,934 | 245,135 | 243,193 |
| R-squared | 0.011 | 0.236 | 0.004 | 0.254 |

| **Panel B: predicting forecast errors by sign of past consensus errors** | | | | |
|---|---|---|---|---|
| Sub-sample: | $\text{Mean}_{i,t}(\text{Forecast Error}_{a,i,t}) < 0$ | | $\text{Mean}_{i,t}(\text{Forecast Error}_{a,i,t}) \geq 0$ | |
| | (1) | (2) | (3) | (4) |
| Dependent variable: | Forecast $\text{Error}_{a,i,t+1}$ | | | |
| $\beta_m$: $\text{Revision}_{a,i,t}$ | 0.788** | 0.766 | 0.287*** | 0.137 |
| | (0.346) | (0.493) | (0.100) | (0.109) |
| $\beta_f$: $\text{Revision}_{a,i,t}$ x $1(\text{Female-led})_{i,t}$ | -0.755** | -0.514** | -0.178 | -0.112 |
| | (0.291) | (0.204) | (0.168) | (0.152) |
| Observations | 177,985 | 175,702 | 307,083 | 304,936 |
| R-squared | 0.016 | 0.198 | 0.001 | 0.167 |

| **Panel C: predicting forecast errors by sign of past abnormal returns** | | | | |
|---|---|---|---|---|
| Sub-sample: | Abnormal $\text{Return}_{i,t} < 0$ | | Abnormal $\text{Return}_{i,t} \geq 0$ | |
| | (1) | (2) | (3) | (4) |
| Dependent variable: | Forecast $\text{Error}_{a,i,t+1}$ | | | |
| $\beta_m$: $\text{Revision}_{a,i,t}$ | 1.002** | 0.660 | 0.289* | 0.344 |
| | (0.463) | (0.419) | (0.146) | (0.204) |
| $\beta_f$: $\text{Revision}_{a,i,t}$ x $1(\text{Female-led})_{i,t}$ | -0.988** | -0.862* | -0.212* | -0.110 |
| | (0.464) | (0.447) | (0.104) | (0.107) |
| Observations | 243,369 | 241,013 | 250,825 | 248,422 |
| R-squared | 0.003 | 0.259 | 0.003 | 0.284 |
| Controls | No | Yes | No | Yes |
| Forecast Period, Analyst FEs | No | Yes | No | Yes |
| Firm, Broker FEs | No | Yes | No | Yes |
| Std Errors (clustered) | Industry&Year | Industry&Year | Industry&Year | Industry&Year |

Notes: Panels A, B, and C present the results from a regression of forecast errors, Forecast $\text{Error}_{a,i,t+1}$, on forecast revisions, $\text{Revision}_{a,i,t}$, an indicator of whether the firm has a female CEO/co-CEO, $1(\text{Female-led})_{i,t}$, and its interaction with forecast revisions. The particular specification is: Forecast $\text{Error}_{a,i,t+1} = \beta_0 + \beta_1 \times 1(\text{Female-led})_{i,t} + \beta_m \times \text{Revision}_{a,i,t} + \beta_f \times \text{Revision}_{a,i,t} \times 1(\text{Female-led})_{i,t} + u_{a,i,t+1}$, where $a$ represents an analyst, $i$, a firm-CEO pair, and $t$, a quarter. Panel A shows results for two sub-samples: negative current forecast errors, Forecast Error(t), in columns 1 and 2, and positive current forecast errors in columns 3 and 4. Panel B displays estimates for the sub-sample of negative consensus error, Mean(Forecast Error(t)), in columns 1 and 2, and of positive consensus error in columns 3 and 4. Panel C presents regressions in sub-samples based on the abnormal return registered by the firm's stock on earnings announcement day: negative abnormal returns in columns 1 and 2, and positive abnormal returns in columns 3 and 4. In all three panels, columns 2 and 4 control for firm-level variables (past forecasts, Forecast(t-2); market capitalization; an indicator of whether capitalization is lower than \$2 billion, and its interaction with forecast revisions; an indicator of whether capitalization is higher than \$10 billion, and its interaction with forecast revisions), and adds a series of fixed effects (forecast period, analyst, firm, and broker). Market capitalization is measured two quarters before the forecast period (i.e., at t-1). Standard errors are clustered at the 4-digit SIC code and year levels. If a firm has more than one CEO in a given period, its observations are weighted by the inverse of the total number of co-CEOs at that period. In all regressions, forecast errors and forecast revisions are normalized by the lagged end-of-quarter price of a firm's stock, $P_{i,t-1}$. Forecast revisions are winsorized at the 1% and 99% levels.

Table A5: Predictive regressions: changing the baseline specification

| | | | |
|---|---|---|---|
| **Panel A: full sample** (col. 1, Panel B, Table 3) | | | |
| Sub-sample: | All Revisions | | |
| Data Treatment: | Baseline | Firms Equal-Wt | Trim. Revisions |
| | (1) | (2) | (3) |
| Dependent variable: | Forecast $\text{Error}_{a,i,t+1}$ | | |
| $\beta_m$: $\text{Revision}_{a,i,t}$ | 0.753** | 0.910*** | 0.343*** |
| | (0.298) | (0.302) | (0.116) |
| $\beta_f$: $\text{Revision}_{a,i,t}$ x $1(\text{Female-led})_{i,t}$ | -0.684** | -0.682** | -0.218* |
| | (0.266) | (0.272) | (0.109) |
| **Panel B: sub-sample of negative forecast revisions** (col. 1, Panel A, Table 4) | | | |
| Sub-sample: | Negative Revisions | | |
| Data Treatment: | Baseline | Firms Equal-Wt | Trim. Revisions |
| | (1) | (2) | (3) |
| Dependent variable: | Forecast $\text{Error}_{a,i,t+1}$ | | |
| $\beta_m$: $\text{Revision}_{a,i,t}$ | 1.072** | 1.246** | 0.481** |
| | (0.460) | (0.480) | (0.208) |
| $\beta_f$: $\text{Revision}_{a,i,t}$ x $1(\text{Female-led})_{i,t}$ | -1.051** | -0.922** | -0.408** |
| | (0.409) | (0.447) | (0.180) |
| **Panel C: sub-sample of positive forecast revisions** (col. 1, Panel B, Table 4) | | | |
| Sub-sample: | Positive Revisions | | |
| Data Treatment: | Baseline | Firms Equal-Wt | Trim. Revisions |
| | (1) | (2) | (3) |
| Dependent variable: | Forecast $\text{Error}_{a,i,t+1}$ | | |
| $\beta_m$: $\text{Revision}_{a,i,t}$ | -0.020 | -0.207 | 0.020 |
| | (0.188) | (0.246) | (0.145) |
| $\beta_f$: $\text{Revision}_{a,i,t}$ x $1(\text{Female-led})_{i,t}$ | 0.052 | -0.209 | 0.329** |
| | (0.208) | (0.312) | (0.154) |
| Controls | No | No | No |
| Forecast Period, Analyst FEs | No | No | No |
| Firm, Broker FEs | No | No | No |
| Std Errors (clustered) | Industry&Year | Industry&Year | Industry&Year |

Notes: Panels A, B, and C present robustness for the results in previous tables. Panel A considers the full sample of forecast revisions, while panel B considers only negative forecast revisions, and panel C, only positive forecast revisions. Column 1 in all panels presents the baseline results in Tables 1 and 2. In the baseline, all observations receive the same weight which implies a form of value-weighting (since larger firms have more analysts covering them). Moreover, in the baseline specification, forecast revisions are winsorized at the 1% and 99% levels. Column 2 in all panels displays results when observations are weighted by the inverse of the total number of analysts associated with a company, so that all companies are equally weighted in the sample. Column 3 in all panels shows results when forecast revisions are trimmed at the 1% and 99% levels instead of winsorized (i.e. the lowest and highest 1% of forecast revisions are removed from the sample). Market capitalization is measured two quarters before the forecast period quarter (i.e., at t-1). Standard errors are clustered at the 4-digit SIC code and year levels. If a firm has more than one CEO in a given period, its observations are weighted by the inverse of the total number of co-CEOs at that period.

Table A6: Predictive regressions after negative revisions: weak versus strong signals

| Sub-sample: | All Revisions | Small Revisions | Large Revisions |
|---|---|---|---|
| | (1) | (2) | (3) |
| Dependent Variable: | Forecast Error$_{a,i,t+1}$ | | |
| $\beta_m$: Revision$_{a,i,t}$ | 1.072** | 0.275** | 1.212** |
| | (0.460) | (0.128) | (0.522) |
| $\beta_f$: Revision$_{a,i,t}$ x 1(Female-led)$_{i,t}$ | -1.051** | 0.251 | -1.209** |
| | (0.409) | (0.275) | (0.466) |
| Observations | 294,916 | 147,456 | 147,460 |
| R-squared | 0.004 | 0.000 | 0.004 |
| Controls | Yes | Yes | Yes |
| Forecast Period, Analyst FEs | Yes | Yes | Yes |
| Firm, Broker FEs | Yes | Yes | Yes |
| Std Errors (clustered) | Industry&Year | Industry&Year | Industry&Year |

Notes: This table repeats the specification in Column 3 of panel A Table 2. Column 1 just replicates the result in Column 3 of panel A Table 2. Column 2 uses only forecast revisions that are smaller (in absolute value) than the median of the distribution. Column 3 includes only forecast revisions that are larger (in absolute value) than the median of the distribution. Standard errors are clustered at the 4-digit SIC code and year levels. If a firm has more than one CEO in a given period, its observations are weighted by the inverse of the total number of co-CEOs at that period. In all regressions, forecast errors and forecast revisions are normalized by the lagged end-of-quarter price of a firm's stock, $P_{i,t-1}$. Forecast revisions are winsorized at the 1% and 99% levels.

Table A7: Predictive regressions: controlling for other firm and CEO characteristics

| **Panel A: sub-sample of negative forecast revisions** (baseline in col. 3, Panel A, Table 4) | | | | |
|---|---|---|---|---|
| Sub-sample: | Negative Revisions | | | |
| Characteristic: | Baseline<br>(1) | CEO Tenure<br>(2) | CEO Age<br>(3) | Past Prices<br>(4) |
| Dependent variable: | Forecast Error$_{a,i,t+1}$ | | | |
| $\beta_m$: Revision$_{a,i,t}$ | 0.812* | 0.857* | 0.921* | 1.060* |
| | (0.463) | (0.458) | (0.491) | (0.579) |
| $\beta_f$: Revision$_{a,i,t}$ x 1(Female-led)$_{i,t}$ | -0.838*** | -0.852*** | -0.837*** | -0.861*** |
| | (0.272) | (0.261) | (0.282) | (0.273) |
| $\beta_{top}$: Revision$_{a,i,t}$ x 1(Top 50%)$_{i,t}$ | | -0.106 | -0.205 | -1.135* |
| | | (0.180) | (0.337) | (0.662) |
| Observations | 292,483 | 292,483 | 288,267 | 292,483 |
| R-squared | 0.278 | 0.278 | 0.278 | 0.278 |
| Controls | Yes | Yes | Yes | Yes |
| Forecast period, Analyst FEs | Yes | Yes | Yes | Yes |
| Firm, Broker FEs | Yes | Yes | Yes | Yes |
| Std Errors (clustered) | Industry&Year | Industry&Year | Industry&Year | Industry&Year |

| **Panel B: sub-sample of positive forecast revisions** (baseline in col. 3, Panel B, Table 4) | | | | |
|---|---|---|---|---|
| Sub-sample: | Positive Revisions | | | |
| Characteristic: | Baseline<br>(1) | CEO Tenure<br>(2) | CEO Age<br>(3) | Past Prices<br>(4) |
| Dependent variable: | Forecast Error$_{a,i,t+1}$ | | | |
| $\beta_m$: Revision$_{a,i,t}$ | 0.302 | 0.191 | 0.214 | 0.319 |
| | (0.264) | (0.202) | (0.244) | (0.298) |
| $\beta_f$: Revision$_{a,i,t}$ x 1(Female-led)$_{i,t}$ | 0.045 | 0.067 | 0.061 | 0.044 |
| | (0.235) | (0.247) | (0.232) | (0.238) |
| $\beta_{top}$: Revision$_{a,i,t}$ x 1(Top 50%)$_{i,t}$ | | 0.242 | 0.155 | -0.092 |
| | | (0.248) | (0.271) | (0.147) |
| Observations | 207,362 | 207,362 | 204,507 | 207,362 |
| R-squared | 0.159 | 0.159 | 0.159 | 0.159 |
| Controls | Yes | Yes | Yes | Yes |
| Forecast Period, Analyst FEs | Yes | Yes | Yes | Yes |
| Firm, Broker FEs | Yes | Yes | Yes | Yes |
| Std Errors (clustered) | Industry&Year | Industry&Year | Industry&Year | Industry&Year |

Notes: Panel A and panel B add additional controls to the regressions in previous tables. Column 1 in all panels presents the baseline results in col. 3 of panel A in Table 2. In the baseline, regressions control for analyst- and firm-level variables (past forecasts, Forecast(t-2); market capitalization; an indicator of whether capitalization is lower than $2 billion, and its interaction with forecast revisions; an indicator of whether capitalization is higher than $10 billion, and its interaction with forecast revisions) and add a series of fixed effects (forecast period fixed effects, analyst fixed effects, firm fixed effects, and fixed effects for the broker an analyst belongs to). Moreover, forecast errors and forecast revisions are normalized by the lagged end-of-quarter price of a firm's stock, $P_{i,t-1}$. In the baseline specification, forecast revisions are also winsorized at the 1% and 99% levels. Columns 2 through 4 extend the baseline specification to control for a different CEO- and firm-level characteristics, controlling for: the level of that characteristic, an indicator of whether the firm's CEO is in the top 50% of the characteristic's distribution, and the interaction of the last indicator with forecast revisions. The particular specification is: $\text{Forecast Error}_{a,i,t+1} = \beta_0 + \beta_1 \times 1(\text{Female-led})_{i,t} + \beta_m \times \text{Revision}_{a,i,t} + \beta_f \times \text{Revision}_{a,i,t} \times 1(\text{Female-led})_{i,t} + \beta_{top} \times \text{Revision}_{a,i,t} \times 1\{\text{Characteristic}_{i,t} \text{ in Top } 50\%\} + \beta_2 \times 1\{\text{Characteristic}_{i,t} \text{ in Top } 50\%\} + u_{a,i,t+1}$, where $a$ is an analyst, $i$ is a firm-CEO pair, and $t$ is a forecast period. The characteristic in column 2 of both panels is CEO tenure, in column 3, CEO age, and in column 4, past prices ($P_{t-1}$). Market capitalization is measured two quarters before the forecast period (i.e., at t-1). Standard errors are clustered at the 4-digit SIC code and year levels. If a firm has more than one CEO in a given period, its observations are weighted by the inverse of the total number of co-CEOs at that period.

Table A8: Predictive regressions: mean consensus-level variables

**Panel A: predicting forecast errors after negative forecast revisions**

| | (1) | (2) | (3) |
|---|---|---|---|
| Dependent Variable: | $\text{Mean}_{i,t+1}(\text{Forecast Error}_{a,i,t+1})$ | | |
| $\beta_m$: $\text{Mean}_{i,t}(\text{Revision}_{a,i,t})$ | 1.505*** | 1.127*** | 1.205** |
| | (0.489) | (0.371) | (0.483) |
| $\beta_f$: $\text{Mean}_{i,t}(\text{Revision}_{a,i,t})$ x $1(\text{Female-led})_{i,t}$ | -1.076** | -0.819*** | -0.814** |
| | (0.451) | (0.282) | (0.305) |
| Observations | 97,308 | 96,210 | 96,210 |
| R-squared | 0.004 | 0.224 | 0.224 |
| Firm, Forecast Period FEs | No | Yes | Yes |
| Controls | No | No | Yes |
| Std Errors (clustered) | Industry&Year | Industry&Year | Industry&Year |

**Panel B: predicting forecast errors after positive forecast revisions**

| | (1) | (2) | (3) |
|---|---|---|---|
| Dependent Variable: | $\text{Mean}_{i,t}(\text{Forecast Error}_{a,i,t+1})$ | | |
| $\beta_m$: $\text{Mean}_{i,t}(\text{Revision}_{a,i,t})$ | 0.260 | 0.259 | 0.288** |
| | (0.201) | (0.155) | (0.129) |
| $\beta_f$: $\text{Mean}_{i,t}(\text{Revision}_{a,i,t})$ x $1(\text{Female-led})_{i,t}$ | -0.513 | -0.468 | -0.468 |
| | (0.374) | (0.306) | (0.306) |
| Observations | 66,532 | 65,334 | 65,334 |
| R-squared | 0.001 | 0.172 | 0.172 |
| Firm, Forecast Period FEs | No | Yes | Yes |
| Controls | No | No | Yes |
| Std Errors (clustered) | Industry&Year | Industry&Year | Industry&Year |

Notes: Panel A presents the results from the same specifications as in panel A of Table 2, while panel B presents the results from specifications as in panel B of Table 2. In both panels A and B, regressions are run at the consensus-level, using the average forecast error and average forecast revision across analysts. Standard errors are clustered at the 4-digit SIC code and year levels. If a firm has more than one CEO in a given period, its observations are weighted by the inverse of the total number of co-CEOs at that period. Forecast errors and forecast revisions at the analyst-level are normalized by the lagged end-of-quarter price of a firm's stock, $P_{i,t-1}$. Forecast revisions are winsorized at the 1% and 99% levels.

Table A9: Predictive regressions: median consensus-level variables

| **Panel A: predicting forecast errors after negative forecast revisions** | | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Dependent Variable: | $\text{Median}_{i,t}(\text{Forecast Error}_{a,i,t+1})$ | | |
| $\beta_m$: $\text{Median}_{i,t}(\text{Revision}_{a,i,t})$ | 1.471*** | 1.168** | 1.302** |
| | (0.474) | (0.435) | (0.631) |
| $\beta_f$: $\text{Median}_{i,t}(\text{Revision}_{a,i,t})$ x $1(\text{Female-led})_{i,t}$ | -1.054** | -0.857** | -0.848** |
| | (0.433) | (0.333) | (0.337) |
| Observations | 96,581 | 95,478 | 95,478 |
| R-squared | 0.004 | 0.236 | 0.236 |
| Firm, Forecast Period FEs | No | Yes | Yes |
| Controls | No | No | Yes |
| Std Errors (clustered) | Industry&Year | Industry&Year | Industry&Year |

| **Panel B: predicting forecast errors after positive forecast revisions** | | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Dependent Variable: | $\text{Median}_{i,t}(\text{Forecast Error}_{a,i,t+1})$ | | |
| $\beta_m$: $\text{Median}_{i,t}(\text{Revision}_{a,i,t})$ | 0.135 | 0.111 | 0.410** |
| | (0.244) | (0.175) | (0.163) |
| $\beta_f$: $\text{Median}_{i,t}(\text{Revision}_{a,i,t})$ x $1(\text{Female-led})_{i,t}$ | -0.356 | -0.351 | -0.337 |
| | (0.386) | (0.279) | (0.283) |
| Observations | 67,259 | 66,063 | 66,063 |
| R-squared | 0.000 | 0.138 | 0.138 |
| Firm, Forecast Period FEs | No | Yes | Yes |
| Controls | No | No | Yes |
| Std Errors (clustered) | Industry&Year | Industry&Year | Industry&Year |

Notes: Panel A presents the results from the same specifications as in panel A of Table 2, while panel B presents the results from specifications as in panel B of Table 2. In both panels A and B, regressions are run at the consensus-level, using the median forecast error and median forecast revision across analysts. Standard errors are clustered at the 4-digit SIC code and year levels. If a firm has more than one CEO in a given period, its observations are weighted by the inverse of the total number of co-CEOs at that period. Forecast errors and forecast revisions at the analyst-level are normalized by the lagged end-of-quarter price of a firm's stock, $P_{i,t-1}$. Forecast revisions are winsorized at the 1% and 99% levels.

## Table A10: Predictive regressions: changes in coefficients over time

| Sub-sample: | Year $\geq$ 2000 | | Year $\geq$ 2015 | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Forecast Error$_{a,i,t+1}$ | | | |
| Revision$_{a,i,t}$ x 1(Revision$_{a,i,t}$ $<$ 0) | 1.022** | 0.755* | -0.097 | -0.049 |
| | (0.485) | (0.436) | (0.153) | (0.119) |
| Revision$_{a,i,t}$ x 1(Revision$_{a,i,t}$ $<$ 0) x 1(Female-led)$_{i,t}$ | -1.024** | -0.906** | -0.323 | -0.403*** |
| | (0.428) | (0.361) | (0.201) | (0.105) |
| Revision$_{a,i,t}$ x 1(Revision$_{a,i,t}$ $\geq$ 0) | -0.040 | 0.081 | 0.174 | 0.045 |
| | (0.195) | (0.121) | (0.251) | (0.089) |
| Revision$_{a,i,t}$ x 1(Revision$_{a,i,t}$ $\geq$ 0) x 1(Female-led)$_{i,t}$ | 0.063 | 0.043 | 0.029 | 0.178 |
| | (0.215) | (0.247) | (0.266) | (0.245) |
| Observations | 472,797 | 471,240 | 187,145 | 186,675 |
| R-squared | 0.003 | 0.134 | 0.001 | 0.232 |
| Controls | No | Yes | No | Yes |
| Forecast Period, Analyst FEs | No | Yes | No | Yes |
| Firm, Broker FEs | No | Yes | No | Yes |
| Cluster Level | Industry&Year | Industry&Year | Industry&Year | Industry&Year |

Notes: This table presents results from the same specifications as in Table 3 for different sub-samples. While Table 3 uses data from 1984 through 2022, columns 1 and 2 show results for the sample going from 2000 to 2022, and columns 3 and 4, from 2015 to 2022. Standard errors are clustered at the 4-digit SIC code and year levels. If a firm has more than one CEO in a given period, its observations are weighted by the inverse of the total number of co-CEOs at that period. Forecast errors and forecast revisions at the analyst-level are normalized by the lagged end-of-quarter price of a firm's stock, $P_{i,t-1}$. Forecast revisions are winsorized at the 1% and 99% levels.

Table A11: Predictive regressions: heterogeneity across characteristics

| Characteristic: | Firm Size (1) | CEO Tenure (2) | CEO Age (3) | Past Prices (4) |
|---|---|---|---|---|
| Dependent variable: | Forecast $\text{Error}_{a,i,t+1}$ | | | |
| $\gamma_{m,top}$: $\text{Revision}_{a,i,t}$ x $1(\text{Top } 50\%)_{i,t}$ | 1.103* | 0.753* | 0.744* | 0.178 |
| | (0.605) | (0.438) | (0.429) | (0.128) |
| $\gamma_{f,top}$: $\text{Revision}_{a,i,t}$ x $1(\text{Top } 50\%)_{i,t}$ x $1(\text{Female-led})_{i,t}$ | -0.980* | -0.585* | -0.780** | -0.013 |
| | (0.532) | (0.301) | (0.362) | (0.225) |
| $\gamma_{m,bottom}$: $\text{Revision}_{a,i,t}$ x $1(\text{Bottom } 50\%)_{i,t}$ | 0.726* | 0.860** | 0.890* | 0.897** |
| | (0.362) | (0.369) | (0.448) | (0.423) |
| $\gamma_{f,bottom}$: $\text{Revision}_{a,i,t}$ x $1(\text{Bottom } 50\%)_{i,t}$ x $1(\text{Female-led})_{i,t}$ | -0.790*** | -0.988*** | -0.905** | -0.967*** |
| | (0.265) | (0.286) | (0.366) | (0.317) |
| Observations | 292,483 | 292,483 | 289,385 | 292,483 |
| R-squared | 0.278 | 0.278 | 0.278 | 0.278 |
| Controls | Yes | Yes | Yes | Yes |
| Forecast Period, Analyst FEs | Yes | Yes | Yes | Yes |
| Firm, Broker FEs | Yes | Yes | Yes | Yes |
| Std Errors (clustered) | Industry&Year | Industry&Year | Industry&Year | Industry&Year |
| $\gamma_{m,top} - \gamma_{m,bottom}$ | 0.378 | -0.108 | -0.147 | -0.719 |
| $H_0$: $\gamma_{m,top} - \gamma_{m,bottom}$ (p-value) | 0.396 | 0.548 | 0.704 | 0.044 |
| $\gamma_{m,top} + \gamma_{f,top} - (\gamma_{m,top} + \gamma_{f,bottom})$ | 0.188 | 0.295 | -0.022 | 0.235 |
| $H_0$: $\gamma_{m,top} + \gamma_{f,top} - (\gamma_{m,top} + \gamma_{f,bottom})$ (p-value) | 0.311 | 0.202 | 0.901 | 0.417 |

Notes: This table expands on the regression in Column 3 of panel A Table 2. In this table, I further interact forecast revisions with whether an observation is in the top-50% or in the bottom 50% of the distribution of characteristics. The particular specification is: Forecast $\text{Error}_{a,i,t+1} = \gamma_0 + \gamma_{m,top} \times \text{Revision}_{a,i,t} \times 1\{\text{Characteristic}_{i,t} \text{ in Top } 50\%\} + \gamma_{f,top} \times \text{Revision}_{a,i,t} \times 1\{\text{Characteristic}_{i,t} \text{ in Top } 50\%\} \times 1(\text{Female-led})_{i,t} + \gamma_{m,bottom} \times \text{Revision}_{a,i,t} \times 1\{\text{Characteristic}_{i,t} \text{ in Bottom } 50\%\} + \gamma_{f,bottom} \times \text{Revision}_{a,i,t} \times 1\{\text{Characteristic}_{i,t} \text{ in Bottom } 50\%\} \times 1(\text{Female-led})_{i,t} + \gamma_1 \times 1\{\text{Characteristic}_{i,t} \text{ in Bottom } 50\%\} + \gamma_2 \times 1(\text{Female-led})_{i,t} + \gamma_3 \times 1\{\text{Characteristic}_{i,t} \text{ in Bottom } 50\%\} \times 1(\text{Female-led})_{i,t} + u_{a,i,t+1}$, where $i$ is a firm-CEO pair, $a$ is an analyst, and $t$ is a forecast period. Standard errors are clustered at the 4-digit SIC code and year levels. If a firm has more than one CEO in a given period, its observations are weighted by the inverse of the total number of co-CEOs at that period. In all regressions, forecast errors and forecast revisions are normalized by the lagged end-of-quarter price of a firm's stock, $P_{i,t-1}$. Forecast revisions are winsorized at the 1% and 99% levels.

Table A12: Predictive regressions: by sign of past yearly growth

**Panel A: analyst-level regressions**

| Dependent Variable: | Forecast $\text{Error}_{a,i,t+1}$ | |
|---|---|---|
| | (1) | (2) |
| $\text{Revision}_{a,i,t}$ | 0.617 | 0.467 |
| | (0.432) | (0.417) |
| $\text{Revision}_{a,i,t} \times 1(\text{YoY Growth}_{i,t}<0)$ | 0.189 | 0.078 |
| | (0.443) | (0.440) |
| Observations | 432,836 | 431,434 |
| R-squared | 0.002 | 0.199 |
| Firm, Forecast Period FEs | No | Yes |
| Analyst, Broker FEs | No | Yes |
| Cluster Level | Industry&Year | Industry&Year |

**Panel B: consensus-level regressions**

| Dependent Variable: | $\text{Mean}_{i,t+1}(\text{Forecast Error}_{a,i,t+1})$ | |
|---|---|---|
| | (1) | (2) |
| $\text{Mean}_{i,t}(\text{Revision}_{a,i,t})$ | 1.164 | 1.125 |
| | (0.790) | (0.839) |
| $\text{Mean}_{i,t}(\text{Revision}_{a,i,t}) \times 1(\text{YoY Growth}_{i,t}<0)$ | -0.107 | -0.268 |
| | (0.725) | (0.818) |
| Observations | 125,132 | 124,780 |
| R-squared | 0.003 | 0.133 |
| Firm, Forecast Period FEs | No | Yes |
| Cluster Level | Industry&Year | Industry&Year |

Notes: Panel A present the results from a regression at the analyst level of future forecast errors, Forecast Error(t+1) (Forecast $\text{Error}_{a,i,t+1}$), on current forecast revisions, Revision(t) ($\text{Revision}_{a,i,t}$), an indicator of whether the firm experienced negative yearly earnings growth, $1(\text{YoY Growth}_{i,t} < 0)$, and its interaction with forecast revisions. The particular specification is: $\text{Forecast Error}_{a,i,t+1} = \beta_0 + \beta_1 \times 1(\text{YoY Growth} < 0)_{i,t} + \beta_p \text{Revision}_{a,i,t} + \beta_n \times \text{Revision}_{a,i,t} \times 1(\text{YoY Growth} < 0)_{i,t} + u_{a,i,t+1}$, where $a$ represents an analyst, $i$, a firm-CEO pair, and $t$, a quarter. Panel B presents the results from a similar regression at the consensus level, where I take the mean of forecast errors and revisions across analysts. The particular specification in these columns is $\text{Mean}_{i,t+1}(\text{Forecast Error}_{a,i,t+1}) = \beta_0 + \beta_1 \times 1(\text{YoY Growth}_{i,t} < 0) + \beta_p \text{Mean}_{i,t}(\text{Revision}_{a,i,t}) + \beta_n \times \text{Mean}_{i,t}(\text{Revision}_{a,i,t}) \times 1(\text{YoY Growth}_{i,t} < 0) + u_{i,t+1}$, where $a$ represents an analyst, $i$, a firm-CEO pair, and $t$, a quarter. Column 2 in Panel A controls for a series of fixed effects: forecast period fixed effects, analyst fixed effects, firm fixed effects, and fixed effects for the broker an analyst belongs to. Column 2 in Panel B controls for firm fixed effects and forecast period fixed effects. Standard errors are clustered at the 4-digit SIC code and year levels. If a firm has more than one CEO in a given period, its observations are weighted by the inverse of the total number of co-CEOs at that period. In all regressions, forecast errors and forecast revisions are normalized by the lagged end-of-quarter price of a firm's stock, $P_{i,t-1}$. Forecast revisions are winsorized at the 1% and 99% levels.

## Table A13: Average forecast errors by CEO gender

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dependent Variable: | Forecast $\text{Error}_{a,i,t+1}$ $\times 100$ | | | |
| $\alpha_f$: 1(Female-led)$_{i,t+1}$ | 0.171** | 0.172** | 0.161* | 0.123 |
| | (0.079) | (0.079) | (0.087) | (0.082) |
| $\alpha_m$: Constant | -0.135 | -0.138 | -0.127*** | -0.019 |
| | (0.089) | (0.090) | (0.001) | (0.037) |
| 1(Revision$_{a,i,t}$ < 0) | | | | -0.185*** |
| | | | | (0.064) |
| 1(Revision$_{a,i,t}$ < 0) x 1(Female-led)$_{i,t}$ | | | | 0.067 |
| | | | | (0.076) |
| Observations | 504,815 | 504,815 | 502,710 | 502,710 |
| R-squared | 0.000 | 0.000 | 0.132 | 0.132 |
| Controls | No | Yes | Yes | Yes |
| Forecast Period, Analyst FEs | No | No | Yes | Yes |
| Firm, Broker FEs | No | No | Yes | Yes |
| Std Errors (clustered) | Industry&Year | Industry&Year | Industry&Year | Industry&Year |
| $\alpha_f + \alpha_m$ | 0.036 | 0.034 | 0.034 | 0.104 |
| $H_0 : \alpha_f + \alpha_m = 0$ (p-value) | 0.605 | 0.631 | 0.698 | 0.222 |

Notes: This table presents the results from the regression Forecast $\text{Error}_{a,i,t+1} = \alpha_m + \alpha_f 1(\text{Female-led})_{i,t} + u_{a,i,t}$, where $i$ is a firm-CEO pair, $a$ an analyst, and $t$ a forecast period. Columns 2 through 4 control for firm-level variables: market capitalization and past forecasts, Forecast(t-1). Columns 3 through 4 include a series of fixed effects: forecast period fixed effects, analyst fixed effects, firm fixed effects, and fixed effects for the broker an analyst belongs to. Column 4 further controls for CEO tenure and its interaction with the female CEO/co-CEO indicator. Market capitalization is measured two quarters before the forecast period (i.e., at t-1). Standard errors are clustered at the 4-digit SIC code and year levels. In all regressions, forecast errors are normalized by the lagged end-of-quarter price of a firm's stock, $P_{i,t-1}$. If a firm has more than one CEO in a given period, its observations are weighted by the inverse of the total number of co-CEOs at that period.

Table A14: Average market surprise by CEO gender

**Panel A: Value-weighted regressions**

| | (1) | (2) |
|---|---|---|
| | $\frac{CS}{PastPrices}_{i,t}$ | $\text{FOM}_{i,t}$ |
| $1(\text{Female-led})_{i,t}$ | 0.005 | -0.017 |
| | (0.013) | (0.032) |
| Constant | 0.069*** | 0.446*** |
| | (0.000) | (0.001) |
| Observations | 229,982 | 229,997 |
| R-squared | 0.236 | 0.285 |
| Firm, Announcement Date FEs | Yes | Yes |
| Std Errors (clustered) | Industry&Year | Industry&Year |

**Panel B: Equally-weighted regressions**

| | (1) | (2) |
|---|---|---|
| | $\frac{CS}{PastPrices}_{i,t}$ | $\text{FOM}_{i,t}$ |
| $1(\text{Female-led})_{i,t}$ | 0.023 | 0.010 |
| | (0.014) | (0.021) |
| Constant | 0.056*** | 0.283*** |
| | (0.000) | (0.001) |
| Observations | 231,597 | 231,732 |
| R-squared | 0.150 | 0.175 |
| Firm, Announcement Date FEs | Yes | Yes |
| Std Errors (clustered) | Industry&Year | Industry&Year |

Notes: This table presents the results from simple regressions at the firm-CEO and announcement date levels of the different measures of surprise on a constant and an indicator for whether the firm has a female CEO/co-CEO. In both panels A and B, if a firm has more than one CEO in a given period, its observations are weighted by the inverse of the total number of co-CEOs at that period. In panel A, observations are moreover weighted by the firm's market capitalization at $t-20$ days. All columns in panel A and in panel B include firm and announcement date fixed effects.

Table A15: Stock market reaction to earnings announcements (value-weighted)

**Panel A: using the FOM score to measure surprises**

| | (1) | (2) | (3) |
|---|---|---|---|
| Dependent Variable: | | Abnormal Return$_{i,t}$ | |
| $\delta_{n,m}$: Surprise$_{i,t}$ x 1(Surprise$_{i,t}$ < 0) | 1.583*** | 1.962*** | 1.944*** |
| | (0.164) | (0.370) | (0.346) |
| $\delta_{n,f}$: Surprise$_{i,t}$ x 1(Surprise$_{i,t}$ < 0) x 1(Female-led)$_{i,t}$ | 0.994** | 0.877** | 0.818* |
| | (0.401) | (0.403) | (0.470) |
| $\delta_{p,m}$: Surprise$_{i,t}$ x 1(Surprise$_{i,t}$ $\geq$ 0) | 1.425*** | 1.373*** | 1.485*** |
| | (0.093) | (0.264) | (0.181) |
| $\delta_{p,f}$: Surprise$_{i,t}$ x 1(Surprise$_{i,t}$ $\geq$ 0) x 1(Female-led)$_{i,t}$ | -0.060 | -0.037 | -0.092 |
| | (0.244) | (0.253) | (0.256) |
| Observations | 232,775 | 229,562 | 228,376 |
| R-squared | 0.041 | 0.043 | 0.157 |
| Firm, Announcement Date FEs | No | Yes | Yes |
| Firm-level Controls | No | Yes | Yes |
| Cluster Level | Industry&Year | Industry&Year | Industry&Year |
| $\delta_{p,m} - \delta_{n,m}$ | -0.158 | -0.589 | -0.459 |
| $H_0: \ \delta_{p,m} - \delta_{n,m} = 0$ (p-value) | 0.331 | 0.188 | 0.197 |
| $\delta_{p,m} + \delta_{p,f} - (\delta_{n,m} + \delta_{n,f})$ | -1.212 | -1.503 | -1.369 |
| $H_0: \ \delta_{p,m} + \delta_{p,f} - (\delta_{n,m} + \delta_{n,f}) = 0$ (p-value) | 0.012 | 0.037 | 0.041 |

**Panel B: using the $CS/PastPrices$ to measure surprises**

| | (1) | (2) | (3) |
|---|---|---|---|
| Dependent Variable: | | Abnormal Return$_{i,t}$ | |
| $\delta_{n,m}$: Surprise$_{i,t}$ x 1(Surprise$_{i,t}$ < 0) | 0.075** | 0.049 | 0.027 |
| | (0.028) | (0.046) | (0.047) |
| $\delta_{n,f}$: Surprise$_{i,t}$ x 1(Surprise$_{i,t}$ < 0) x 1(Female-led)$_{i,t}$ | 0.286** | 0.225* | 0.200 |
| | (0.115) | (0.130) | (0.130) |
| $\delta_{p,m}$: Surprise$_{i,t}$ x 1(Surprise$_{i,t}$ $\geq$ 0) | 0.287*** | 0.263** | 0.457*** |
| | (0.056) | (0.103) | (0.135) |
| $\delta_{p,f}$: Surprise$_{i,t}$ x 1(Surprise$_{i,t}$ $\geq$ 0) x 1(Female-led)$_{i,t}$ | 0.093 | 0.049 | 0.054 |
| | (0.143) | (0.123) | (0.120) |
| Observations | 232,755 | 229,559 | 228,373 |
| R-squared | 0.031 | 0.031 | 0.146 |
| Firm, Announcement Date FEs | No | Yes | Yes |
| Firm-level Controls | No | Yes | Yes |
| Cluster Level | Industry&Year | Industry&Year | Industry&Year |
| $\delta_{p,m} - \delta_{n,m}$ | 0.212 | 0.214 | 0.430 |
| $H_0: \ \delta_{p,m} - \delta_{n,m} = 0$ (p-value) | 0.001 | 0.026 | 0.006 |
| $\delta_{p,m} + \delta_{p,f} - (\delta_{n,m} + \delta_{n,f})$ | 0.019 | 0.039 | 0.284 |
| $H_0: \ \delta_{p,m} + \delta_{p,f} - (\delta_{n,m} + \delta_{n,f}) = 0$ (p-value) | 0.924 | 0.768 | 0.244 |

Notes: In both panels A and B, the dependent variable in regressions is the abnormal return (AR) over the S&P 500 index on earnings announcement day. In both panels, the regression specification is: $\text{AR}_{i,t} = \delta_0 + \delta_{m,p} \times 1\{\text{Surprise}_{i,t} \geq 0\} \times \text{Surprise}_{i,t} + \delta_{f,p} \times 1\{\text{Surprise}_{i,t} \geq 0\} \times \text{Surprise}_{i,t} \times 1(\text{Female-led})_{i,t} + \delta_{m,n} \times 1\{\text{Surprise}_{i,t} < 0\} \times \text{Surprise}_{i,t} + \delta_{f,n} \times 1\{\text{Surprise}_{i,t} < 0\} \times \text{Surprise}_{i,t} \times 1(\text{Female-led})_{i,t} + \delta_1 \times 1\{\text{Surprise}_{i,t} < 0\} + \delta_2 \times 1(\text{Female-led})_{i,t} + \delta_3 \times 1\{\text{Surprise}_{i,t} < 0\} \times 1(\text{Female-led})_{i,t} + u_{i,t}$, where $i$ is a firm-CEO pair and $t$, an announcement date. The measure of surprise in panel A is the FOM score, following the methodology in Chiang, Dai, Fan, Hong, and Tu (2019), while the measure of surprise in panel B is the average surprise across analysts (consensus surprise) normalized by the firm's stock price at $t - 20$ days. In both panels, column 2 controls for: the firm's market capitalization in $t - 20$ days, indicators for whether market capitalization is lower than \$2 billion and whether it is higher than \$10 billion, and their interactions with surprises. Column 3 further controls for announcement date fixed effects and firm fixed effects. Standard errors are clustered at the 4-digit SIC code and year of announcement levels. If a firm has more than one CEO in a given period, its observations are weighted by the inverse of the total number of co-CEOs at that period. Observations are moreover weighted by the firm's market capitalization at $t - 20$ days.

Table A16: Market reaction to earnings announcements: changes in coefficients over time

**Panel A: using the FOM score to measure surprises**

| Sub-sample: | Year ≥ 2000 | | Year ≥ 2015 | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Abnormal Return$_{i,t}$ | | | |
| Surprise$_{i,t}$ x 1(Surprise$_{i,t}$ < 0) x 1(Female-led)$_{i,t}$ | 1.164** | 1.234** | 1.390* | 1.179** |
| | (0.517) | (0.538) | (0.628) | (0.488) |
| Surprise$_{i,t}$ x 1(Surprise$_{i,t}$ < 0) | 1.747*** | 1.565*** | 1.700*** | 1.374** |
| | (0.141) | (0.217) | (0.205) | (0.410) |
| Surprise$_{i,t}$ x 1(Surprise$_{i,t}$ ≥ 0) x 1(Female-led)$_{i,t}$ | 0.013 | -0.011 | 0.550 | 0.610 |
| | (0.260) | (0.221) | (0.340) | (0.431) |
| Surprise$_{i,t}$ x 1(Surprise$_{i,t}$ ≥ 0) | 1.521*** | 0.972*** | 1.469*** | 0.714 |
| | (0.084) | (0.177) | (0.128) | (0.567) |
| Observations | 189,989 | 186,754 | 60,948 | 60,162 |
| R-squared | 0.039 | 0.143 | 0.037 | 0.172 |
| Firm, Announcement Date FEs | No | Yes | No | Yes |
| Firm-level Controls | No | Yes | No | Yes |
| Cluster Level | Industry&Year | Industry&Year | Industry&Year | Industry&Year |

**Panel B: using the $CS/PastPrices$ to measure surprises**

| Sub-sample: | Year ≥ 2000 | | Year ≥ 2015 | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Abnormal Return$_{i,t}$ | | | |
| Surprise$_{i,t}$ x 1(Surprise$_{i,t}$ < 0) x 1(Female-led)$_{i,t}$ | 0.169** | 0.195* | 0.174** | 0.266* |
| | (0.072) | (0.094) | (0.066) | (0.115) |
| Surprise$_{i,t}$ x 1(Surprise$_{i,t}$ < 0) | 0.072*** | 0.022 | 0.065 | 0.002 |
| | (0.023) | (0.038) | (0.048) | (0.075) |
| Surprise$_{i,t}$ x 1(Surprise$_{i,t}$ ≥ 0) x 1(Female-led)$_{i,t}$ | 0.008 | -0.074 | 0.139 | -0.014 |
| | (0.116) | (0.100) | (0.167) | (0.139) |
| Surprise$_{i,t}$ x 1(Surprise$_{i,t}$ ≥ 0) | 0.248*** | 0.218* | 0.289*** | 0.470** |
| | (0.048) | (0.106) | (0.065) | (0.142) |
| Observations | 189,970 | 186,751 | 60,943 | 60,161 |
| R-squared | 0.033 | 0.136 | 0.033 | 0.168 |
| Firm, Announcement Date FEs | No | Yes | No | Yes |
| Firm-level Controls | No | Yes | No | Yes |
| Cluster Level | Industry&Year | Industry&Year | Industry&Year | Industry&Year |

Notes: In both panels A and B, the dependent variable in regressions is the abnormal return (AR) over the S&P 500 index on earnings announcement day. In both panels, the regression specification is: $\text{AR}_{i,t} = \delta_0 + \delta_{m,p} \times 1\{\text{Surprise}_{i,t} \geq 0\} \times \text{Surprise}_{i,t} + \delta_{f,p} \times 1\{\text{Surprise}_{i,t} \geq 0\} \times \text{Surprise}_{i,t} \times 1(\text{Female-led})_{i,t} + \delta_{m,n} \times 1\{\text{Surprise}_{i,t} < 0\} \times \text{Surprise}_{i,t} + \delta_{f,n} \times 1\{\text{Surprise}_{i,t} < 0\} \times \text{Surprise}_{i,t} \times 1(\text{Female-led})_{i,t} + \delta_1 \times 1\{\text{Surprise}_{i,t} < 0\} + \delta_2 \times 1(\text{Female-led})_{i,t} + \delta_3 \times 1\{\text{Surprise}_{i,t} < 0\} \times 1(\text{Female-led})_{i,t} + u_{i,t}$, where $i$ is a firm-CEO pair and $t$, an announcement date. The measure of surprise in panel A is the FOM score, following the methodology in Chiang, Dai, Fan, Hong, and Tu (2019), while the measure of surprise in panel B is the average surprise across analysts (consensus surprise) normalized by the firm's stock price at $t - 20$ days. In both panels, columns 2 and 4 control for: the firm's market capitalization in $t - 20$ days, indicators for whether market capitalization is lower than $2 billion and whether it is higher than $10 billion, and their interactions with surprises. Columns 2 and 4 also control for announcement date fixed effects and firm fixed effects. In both panels, columns 1 and 2 consider data from 2000 through 2022, while columns 3 and 4 only consider data from 2015 through 2022. Standard errors are clustered at the 4-digit SIC code and year of announcement levels. If a firm has more than one CEO in a given period, its observations are weighted by the inverse of the total number of co-CEOs at that period. Observations are moreover weighted by the firm's market capitalization at $t - 20$ days.

## Table A17: Annual market-to-book ratios

| Data Treatment: | Equal-weighted | | | Value-weighted | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Dependent Variable: | Annual Market-to-Book Ratio$_{i,t}$ | | | | | |
| 1(Female-led)$_{i,t}$ | 0.528** | 0.510 | 0.534 | -0.127 | -0.593 | -0.447 |
| | (0.231) | (0.313) | (0.385) | (0.745) | (0.817) | (1.142) |
| 1(Surprise$_{i,t}$ $< 0$) | | | -0.340*** | | | 0.227 |
| | | | (0.075) | | | (0.167) |
| 1(Female-led)$_{i,t}$ x 1(Surprise$_{i,t}$ $< 0$) | | | -0.061 | | | -0.359 |
| | | | (0.300) | | | (1.177) |
| Observations | 37,142 | 34,922 | 34,922 | 35,643 | 33,529 | 33,529 |
| R-squared | 0.000 | 0.547 | 0.548 | 0.000 | 0.709 | 0.709 |
| Firm, Year FEs | No | Yes | Yes | No | Yes | Yes |
| Firm-level Controls | No | Yes | Yes | No | Yes | Yes |
| Std Errors (clustered) | Industry&Year | Industry&Year | Industry&Year | Industry&Year | Industry&Year | Industry&Year |
| Mean | 3.559*** | 3.547*** | 3.737*** | 5.035*** | 5.045*** | 4.920*** |
| Standard Error | (0.225) | (0.011) | (0.042) | (0.556) | (0.101) | (0.141) |

Notes: This table shows results of a regression of annual market-to-book ratios on a constant and an indicator of whether the company has a female CEO/co-CEO. Regressions are at the firm-CEO and year levels. Columns 1 through 3 show results for equally weighted regressions, while columns 4 through 6 display results for value-weighted regressions. In this latter case, observations are weighted by the firm's market capitalization in $t - 20$ days. Columns 2 and 3, and 5 and 6 control for CEO tenure and CEO age, firm fixed effects, and year fixed effects. Columns 3 and 6 further control for an indicator variable of whether the company displayed any negative quarterly surprise over the year, and its interaction with the female indicator. Standard errors are clustered at the 4-digit SIC code and year levels. If a firm has more than one CEO in a given period, its observations are weighted by the inverse of the total number of co-CEOs at that period. In columns 4 through 6, observations are moreover weighted by the firm's market capitalization at $t - 20$ days.

Table A18: Average sentiment and disagreement

| Sub-sample: | Analysts | Executives | - |
|---|---|---|---|
| | (1) | (2) | (3) |
| Dependent Variable: | Sentiment$_{i,t}$ | | Sentiment Gap$_{i,t}$ |
| Constant | -0.315*** | 0.943*** | 1.335*** |
| | (0.043) | (0.056) | (0.027) |
| Observations | 77,243 | 77,243 | 77,243 |
| R-squared | 0.000 | 0.000 | 0.000 |
| Firm, Announcement Date FEs | No | No | No |
| Std Errors (clustered) | Industry&Year | Industry&Year | Industry&Year |

Notes: This table presents the results from simple regressions at the firm-earnings-announcement level of the sentiment score or the sentiment gap on a constant. In column 1 the dependent variable is the sentiment score of analysts, computed with the text of the Q&A session of an earnings conference call. In column 2 the dependent variable is the sentiment score of executives calculated based on the text of the presentations part of the earnings conference call. Finally, in column 3 the dependent variable is a measure of disagreement: the sentiment gap — that the absolute value of the difference between analysts' and executives' sentiment scores. Observations are equally weighted. Standard errors are clustered at the 4-digit SIC code and year levels.

Table A19: Disagreement: changes in coefficients after 2015

| Sub-sample: | All Analysts | Male Analysts | Female Analysts |
|---|---|---|---|
| | (1) | (2) | (3) |
| | | Sentiment $\text{Gap}_{i,t}$ | |
| $1(\text{Surprise}_{i,t} < 0)$ x $1(\text{Female-led})_{i,t}$ | 0.086** | 0.108** | 0.034 |
| | (0.027) | (0.045) | (0.083) |
| $1(\text{Surprise}_{i,t} < 0)$ | -0.064*** | -0.061*** | -0.054 |
| | (0.010) | (0.010) | (0.029) |
| $1(\text{Female-led})_{i,t}$ | -0.021 | -0.036 | -0.101 |
| | (0.056) | (0.068) | (0.093) |
| Constant | 1.344*** | 1.346*** | 1.722*** |
| | (0.004) | (0.004) | (0.007) |
| Observations | 38,388 | 38,267 | 17,177 |
| R-squared | 0.387 | 0.369 | 0.271 |
| Firm, Announcement Date FEs | Yes | Yes | Yes |
| Std Errors (clustered) | Industry&Year | Industry&Year | Industry&Year |

Notes: This table presents the results from regressions of the sentiment gap on a constant, an indicator for whether the firm is female-led, and indicator for whether there was a negative earnings surprise on the announcement day, and their interaction. In column 1 the dependent variable is the sentiment gap computed with the questions of all analysts. In column 2 the dependent variable is the sentiment gap calculated based on the questions of male analysts. In column 2 the dependent variable is the sentiment gap calculated based on the questions of female analysts. The sentiment gap is determined by the absolute value of the difference between analysts' and executives' sentiment scores. Observations are equally weighted. Standard errors are clustered at the 4-digit SIC code and year levels.

# B. Appendix Figures

```
Q1 2017 UMH Properties Inc Earnings Call

FREEHOLD May 11, 2017 (Thomson StreetEvents) -- Edited Transcript of UMH Properties Inc earnings conference call or presentation Wednesday, May 10, 2017 at 2:00:00pm GMT

TEXT version of Transcript

==============================================================================
Corporate Participants
==============================================================================
  * Anna T. Chew
    UMH Properties, Inc. - CFO, CAO, VP and Director
  * Brett Taft
    UMH Properties, Inc. - Corporate Officer and VP
  * Eugene W. Landy
    UMH Properties, Inc. - Founder and Chairman
  * Nelli Madden
    UMH Properties, Inc. - Director of Investor Relations
  * Samuel A. Landy
    UMH Properties, Inc. - CEO, President and Director

==============================================================================
Conference Call Participants
==============================================================================
  * Brian Hollenden
    Sidoti &amp; Company, LLC - Research Analyst
  * Craig Kucera
    Wunderlich Securities Inc., Research Division - SVP
  * Michael Boulegeris
    Boulegeris Investments - Investor
  * Paula Poskon
    Stove Advisory Services
```

Figure B1: List of participants in an earnings conference call

Notes: Example of an earnings conference call in Refinitiv.

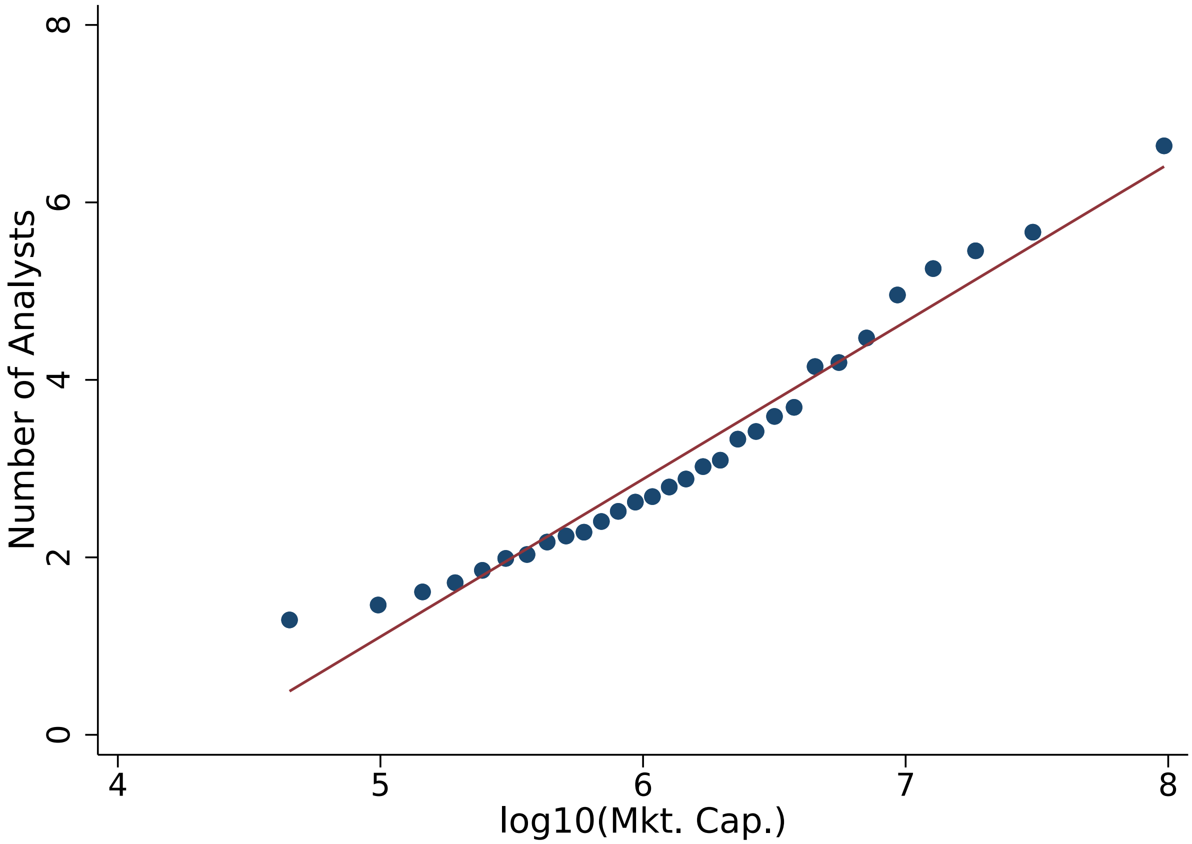
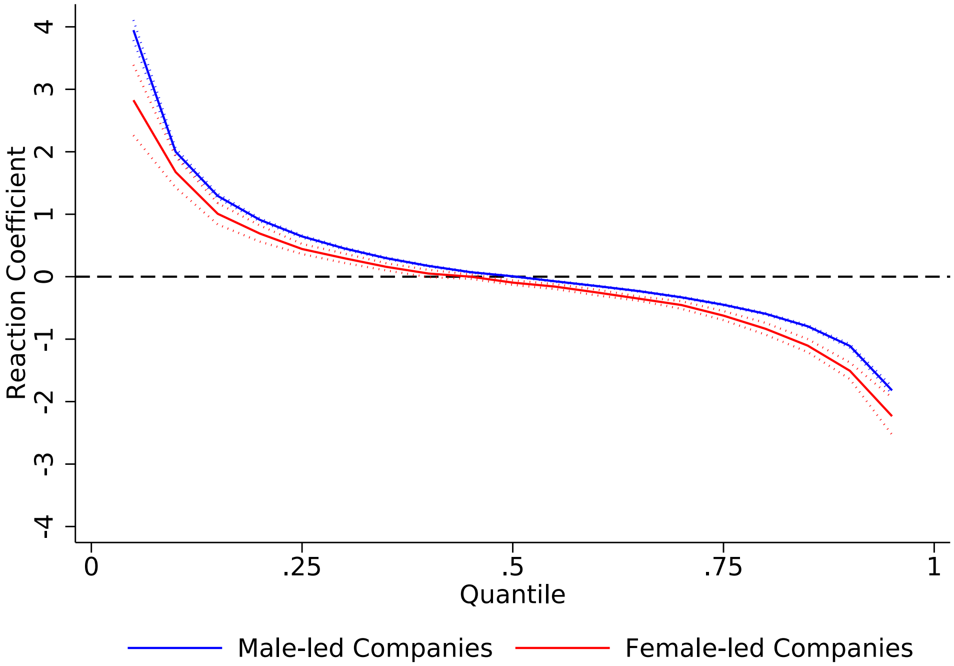
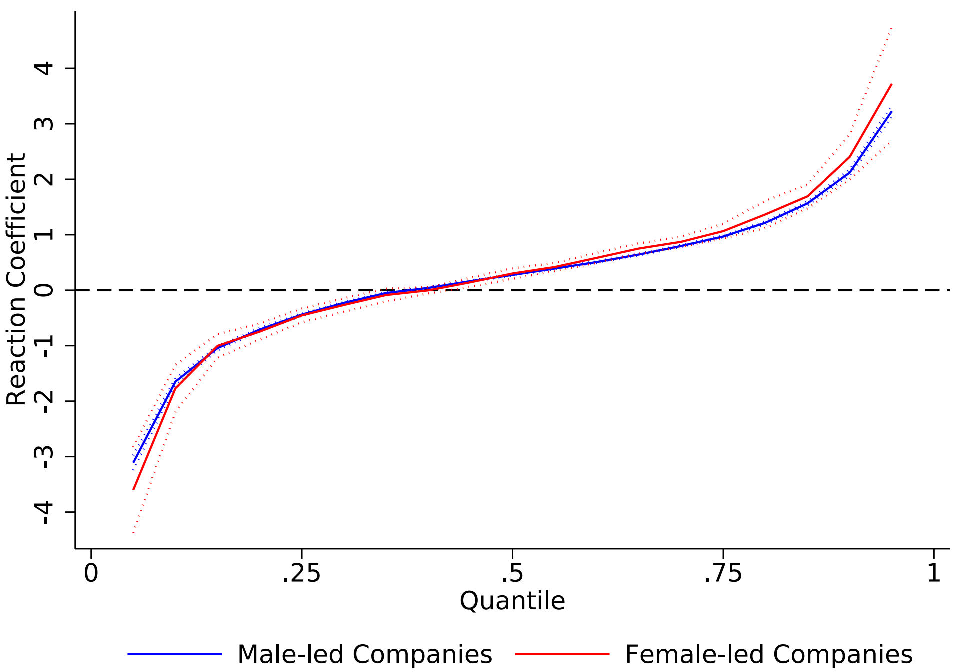



Figure B2: Binned scatter plot: number of analysts versus firm size

Notes: This figure shows a binned scatter plot of the number of analysts that made forecasts for a given company at a given forecast period — say $t$ — against the $\log_{10}$(Market Capitalization) measured at the previous quarter — $t-1$. The figure contains 30 bins. The red line shows the fit of a regression of the number of analysts on the $\log_{10}$(Market Capitalization). The R-squared is of 0.161, and the estimated coefficient is of 1.775 (s.e. 0.120, $p < 0.001$).




(a) Negative Forecast Revisions




(b) Positive Forecast Revisions

Figure B3: Quantile regressions by sign of forecast revisions and CEO gender

Notes: Panels (a) and (b) shows the results from quantile regressions based on the following specification: $fe_{a,i,t} = b_0 + b_1 \cdot 1(Female)_{i,t} + b_2 \cdot \text{Revision}_{a,i,t} + b_3 \cdot \text{Revision}_{a,i,t} \cdot 1(Female)_{i,t}$, where $a$ represents an analyst, $i$, a firm, and $t$ a quarter, and $1(Female)_{i,t}$ indicates if firm $i$ at period $t$ has a female CEO/co-CEO. Panel (a) shows results for the sub-sample of negative forecast revisions, while panel (b) displays results for the sub-sample of positive forecast revisions. In blue, the panels show the point estimate and respective 95% confidence interval for $b_2$ when the regression targets the quantile of forecast errors marked on the x-axis. In red, the panels show the point estimate and respective 95% confidence interval for $b_2 + b_3$.

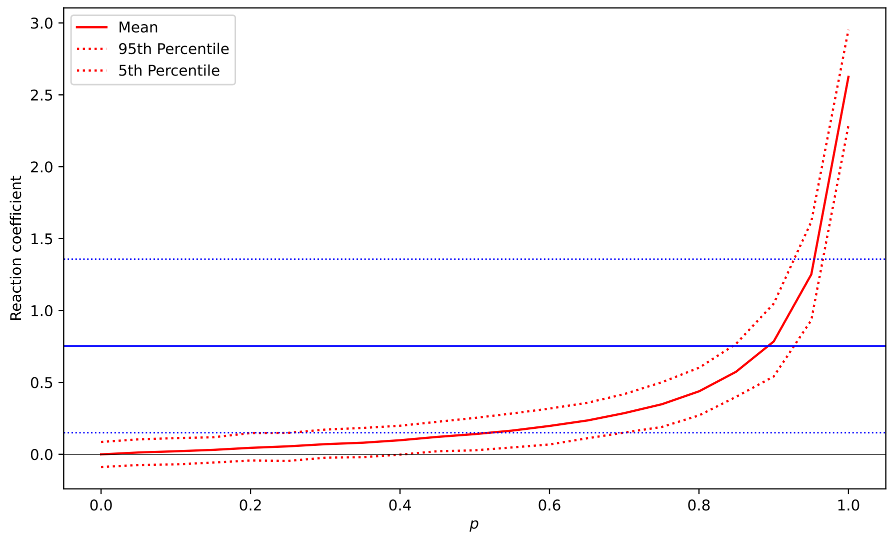



Figure B4: Simulating the model of section 3.3: different probabilities of the bad state of the world

Notes: This panel shows the results of 1,000 simulations of the model in section 3.3 with 1,000 observations each. Each simulation considers a different probability of the bad state of the world: each simulation considers a different $p$. The simulations are based on the following parametrization: $\hat{\sigma}_{\eta,\theta_{bad}} = 2.5 \cdot \sigma_\eta$, $\mu_{bad} = 0$, $\mu_{good} = 0.144$, and $\sigma_v^2 = \sigma_\eta^2 = 0.125/2$. The panel shows the mean and 90% confidence interval for the reaction coefficient of a regression of forecast errors on forecast revisions. The blue lines in panel (a) marks the point estimate and respective 95% confidence interval for the reaction coefficient found in the data (column 1 of panel A in Table 1).

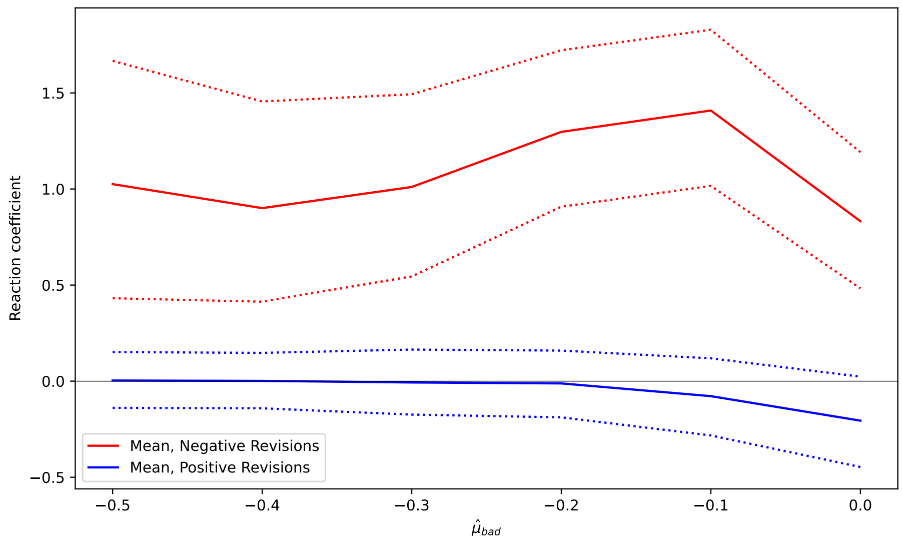
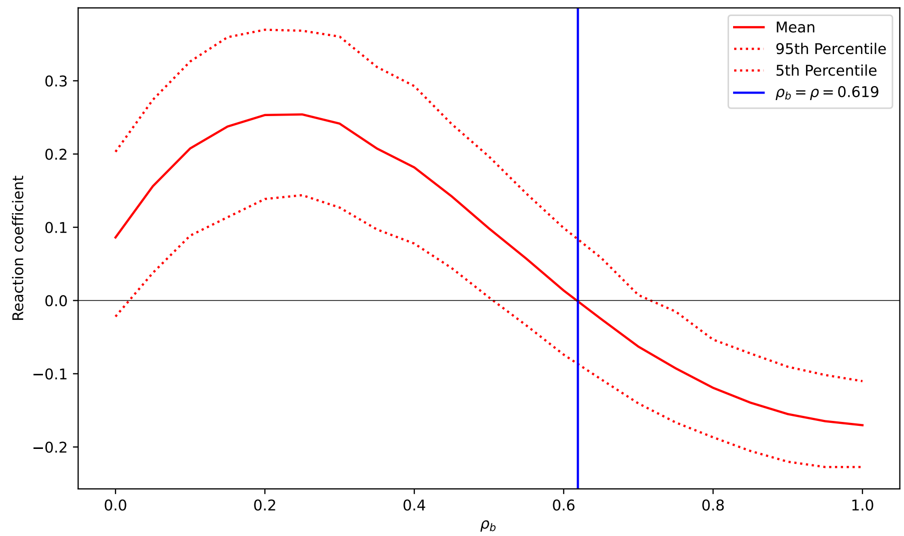
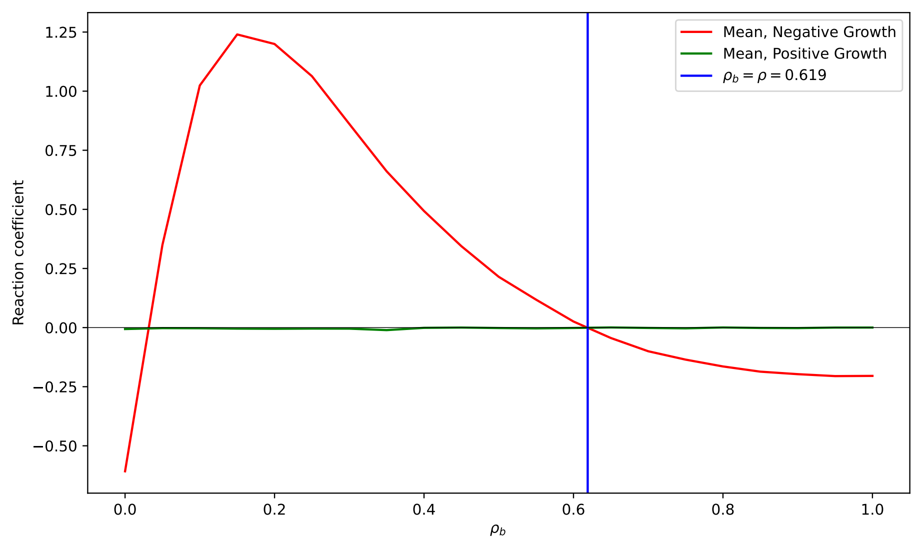



Figure B5: Simulating the model of section 3.3: different values for the perceived mean of fundamental value in the bad state of the world

Notes: This panel shows the results of 1,000 simulations of the model in section 3.3 with 1,000 observations each. Each simulation considers a different value for the perceived mean fundamental value in the bad state of the world: each simulation considers a different $\hat{\mu}_{bad}$. The simulations are based on the following parametrization: $p = 0.5$, $\hat{\sigma}_{\eta,\theta_{bad}} = 2.5 \cdot \sigma_\eta$, $\mu_{bad} = 0$, $\mu_{good} = 0.144$, and $\sigma_v^2 = \sigma_\eta^2 = 0.125/2$. The panel shows, in red, the mean and 90% confidence interval for the reaction coefficient of a regression of forecast errors on forecast revisions after negative forecast revisions. The panel shows, in blue, the mean and 90% confidence interval for the reaction coefficient of a regression of forecast errors on forecast revisions after positive forecast revisions.




(a) Reaction coefficient and confidence interval

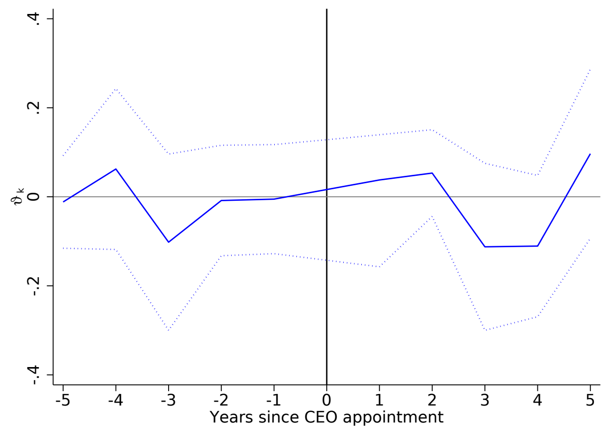



(b) Reaction coefficient by sign of yearly growth

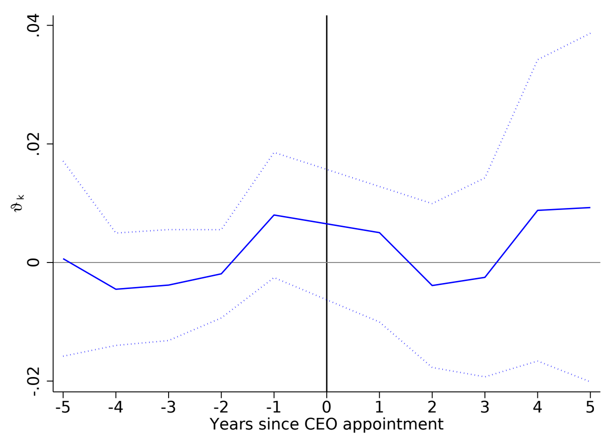
Figure B6: Simulating the model of under-estimation of persistence

Notes: Panels (a) and (b) show the results of 1,000 simulations of the model in section E for each possible value of $\rho_b$ and with 1,000 observations each. The simulations consider a value $\rho = 0.745$ following the estimation in column 3 of panel A in Table A3. Panel (a) shows the mean and 90% confidence interval for the reaction coefficient of a regression of forecast errors on forecast revisions. Panel (b) shows the mean for the reaction coefficient of a regression of forecast errors on forecast revisions conditional on the sign of yearly growth — the green line shows the results of regressions after positive growth, while the red line shows the results of regressions after negative growth.


(a) EPS Growth Rate (QoQ)



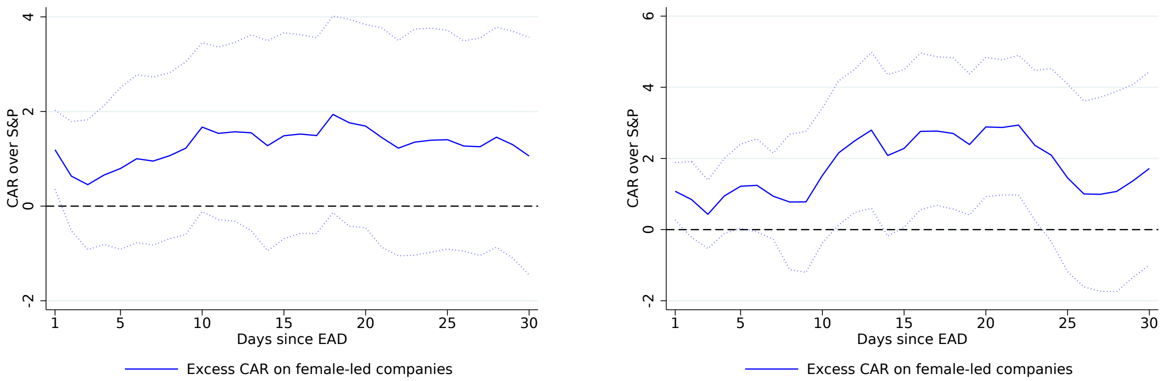

(b) EPS Growth Rate (YoY)




(c) $\frac{EPS}{price}$

Figure B7: Firm conditions around CEO appointment

Notes: The blue lines display coefficients $\vartheta_{k,f}$ estimated and respective 95% interval from a regression following the specification: $Y_{i,t} = \vartheta_{cons} + \sum_{k=-5}^{5} \vartheta_{k,m} \times 1(\text{years since } T = k)_{i,t} + \sum_{k=-5}^{5} \vartheta_{k,f} \times 1(\text{male-to-female transition at } T)_{i,t} \times 1(\text{years since } T = k)_{i,t} + +\vartheta_z \times Z_{i,t} + u_{i,t},$, where $i$ is a firm-CEO pair and $t$ a quarter. $Y_{i,t}$ is either the company's EPS/price (panel (c)), EPS YoY (panel (b)) or QoQ (panel (a)) growth rate. The regressions include firm and quarter fixed effects. If a firm has more than one CEO in a given period, its observations are weighted by the inverse of the total number of co-CEOs at that period.




(a) Equal-weighted (b) Value-weighted

Figure B8: Excess cumulative abnormal return on female-led companies after negative surprises

Notes: The underlying regressions in these graphs follow the specification in column 3 of panel A in Table 4. The particular estimated specification is: $\text{CAR}_{\tau,i,t} = \delta_0 + \delta_{m,p} \times 1\{\text{Surprise}_{i,t} \geq 0\} \times \text{Surprise}_{i,t} + \delta_{f,p} \times 1\{\text{Surprise}_{i,t} \geq 0\} \times \text{Surprise}_{i,t} \times 1(\text{Female-led})_{i,t} + \delta_{m,n} \times 1\{\text{Surprise}_{i,t} < 0\} \times \text{Surprise}_{i,t} + \delta_{f,n} \times 1\{\text{Surprise}_{i,t} < 0\} \times \text{Surprise}_{i,t} \times 1(\text{Female-led})_{i,t} + \delta_1 \times 1\{\text{Surprise}_{i,t} < 0\} + \delta_2 \times 1(\text{Female-led})_{i,t} + \delta_3 \times 1\{\text{Surprise}_{i,t} < 0\} \times 1(\text{Female-led})_{i,t} + u_{i,t}$, where $i$ is a firm-CEO pair and $t$ an announcement date. The blue lines display the estimated coefficient on the interaction between negative surprises and the indicator for whether the company has a female CEO/co-CEO, that is $\delta_{f,n}$. The underlying regression associated with panel (a) is equal-weighted, while the underlying regression associated with panel (b) is value-weighted. In all regressions, if a firm has more than one CEO in a given period, its observations are weighted by the inverse of the total number of co-CEOs at that period.

## C. Determining CEO gender and linking databases

### C.1. Determining the gender of professionals in WRDS Professional

WRDS Professional is a database maintained by S&P Market Intelligence inside the Wharton Research Data Services (WRDS) platform with background information on professionals from companies contained in Capital IQ. The data includes information not only on professionals' full name, but also on the different functions performed by these professionals and their respective time periods (beginning and end dates). Importantly, the database includes a "professional id" that links an executive-firm pair over time, as well as a "person id" that tracks an executive through time and across companies. I also have access to a seniority rank that ranks the different functions executives perform in each company.

In order to determine the gender of an executive, I first extract her given name from her full name. Full name entries in WRDS Professional are stored in the following pattern "'surname', 'given name' 'middle name"'.[24] In the case of composed names, such as 'Jean Paul', I further separate a 'given name' into 'first name' — 'Jean' — and a 'second name' — 'Paul'.

The second step is to attribute gender to given names. I do so by using the following algorithm:

1. Use the "gender detector" package in Python on first names. If no gender is assigned, use the "gender detector" on second names.

2. If no gender has been assigned from 1, use a US dictionary[25] of names to gender on first names. If no gender is assigned, use the dictionary on second names.

3. If no gender has been assigned from 1-2, use a UK dictionary of names to gender.[26] If no gender is assigned, use the dictionary on second names.

[24] In some cases, full names contain a professional's title, such as "PhD" or "MD". I remove these titles from full names.

[25] Available at: https://github.com/OpenGenderTracking/globalnamedata.

[26] Available at: https://github.com/OpenGenderTracking/globalnamedata.

4. If no gender has been assigned from 1-3, make manual fixes.[27] Some WRDS name entries are incomplete, for example containing a first name 'Willia' instead of 'William', or 'Bartholom' instead of 'Bartholomew'. Another important fix is with respect to the first name 'Jean'. This is a female first name according to the algorithm. Nonetheless, when in a composed name — e.g., 'Jean-Paul' —, this is typically a male name. Therefore, I force all instances of 'Jean' in composed names to be male.

This procedure yields gender for 72% of the entries in WRDS Professional, with 15% of them being assigned to female.

### C.2. Determining the gender of speakers in quarterly earnings calls

Files for earnings calls transcripts from Refinitiv can be divided into three parts: (i) information on speakers, (ii) the actual transcript of the call, and (iii) information on the company and call. In order to determine the gender of speakers, the relevant part of the transcript is part (i).

As Figure B1 shows, the heading of the transcript lists corporate participants and outside participants. For each entry in the list,[28] I remove unknown characters and split the entry into the person's: (i) given name, middle name and surname, (ii) position in the company, and (iii) suffix and/or prefix (e.g. 'Mr' or 'Jr'). As with WRDS entries, I split composed names into first and second names.

[27] Here is the full list. Names assigned to male: 'agnaldo', 'andelaney', 'adalmario', 'alondro', 'charle', 'devender', 'edvaldo', 'frabrizio', 'franois', 'filippe', 'heverton', 'jacque', 'motoya', 'niccolo', 'nicols', 'orivaldo', 'rodgrigo', 'sandoval', 'zhenbo', 'johannson', 'rono', 'sloan', 'santanu', 'selby', 'marijn', 'lizst', 'berdon', 'tobey', 'fil', 'natthaniel', 'hap', 'gunan', 'owsley', 'pen', 'gibbons', 'pedros', 'bred', 'pehong', 'sanju', 'gord', 'mohandas', 'balu', 'schond', 'laurans', 'landis', 'cloyce', 'marmion', 'carman', 'cliffe', 'sasson', 'chane', 'dinyar', 'geaton', 'harriss', 'kevi', 'agit', 'leicle', 'basab', 'sabi', 'hollings', 'reay', 'rejean', 'chuan', 'ramalinga', 'stephane', 'fredericus', 'pierric', 'peyton', 'joth', 'jowdat', 'phupinder', 'liecle', 'gordie', 'laxman', 'bohn', 'mariner', 'carey', 'casey', 'ruediger', 'persio', 'joerg', 'kley', 'raf', 'manoe', 'bartholom', 'davi', 'pierr', 'gonzal', 'maurici','bernar', 'reymun', 'rolim', 'nunez', 'alejand', 'augus', 'migue', 'rafae', 'gonzalez', 'joaq', 'pabl', 'mauric', 'thierr', 'nunes', 'reynoso', 'charl', 'orlnado', 'perciv', 'guiller', 'roge', 'victo', 'santiag', 'olievie', 'adema', 'edso', 'osval', 'joaqui', 'edoard', 'aecio', 'affonso', 'affon', 'afranio','agberto', 'agenor','agostin', 'agustino', 'agustinus', 'albertito', 'albertini', 'albilio', 'albinio', 'aleixandre', 'alelio', 'andrze', 'antônio', 'aparecido', 'arthu', 'artu', 'bartholomaus', 'bartholomeu', 'bartomeu', 'benjami', 'andrea', 'patrice', 'willia', 'nickol'. Names assigned to female: 'paulenne', 'viktoriia', 'kerrii', 'dorey', 'nabanita', 'brinlea', 'orsula', 'priscillah', 'nathlie', 'tayne', 'begonya', 'july', 'adriannie', 'agatta', 'agelica', 'angela', 'aparecida'.

[28] Some older transcripts (e.g. from 2002) have these lists mixed together. In these cases, I identify corporate and outside participants by checking the company name associated with them. If the company assigned to them is the same one holding the call, then I consider this speaker to be a corporate participant.

In order to attribute gender, I follow a similar procedure as to that of the last subsection:

1. Use the "gender detector" package in Python on first names. If no gender is assigned, use the "gender detector" on second names.

2. If no gender has been assigned from 1, use a US dictionary of names to gender on first names. If no gender is assigned, use the dictionary on second names.

3. If no gender has been assigned from 1-2, use a UK dictionary of names to gender. If no gender is assigned, use the dictionary on second names.

4. If no gender has been assigned from 1-3, make manual fixes.[29] One important fix is with respect to the first name 'Jean'. This is a female first name according to the algorithm. Nonetheless, when in a composed name — e.g., 'Jean-Paul' —, this is typically a male name. Therefore, I force all instances of 'Jean' in composed names to be male.

5. If no gender has been assigned from 1-4, use information from suffix or prefix. For example, 'Jr' and 'Mr' are assigned to male, while 'Mrs' is assigned to female.

This procedure yields a gender for 97% of the speakers entries in earnings calls from 2001 to 2022, with 12% of them being assigned to female and 85% to male.

## C.3. Matching executives in WRDS Professional, Execucomp, and quarterly earnings calls

After attributing gender to all entries in WRDS Professional and to all corporate participants in earnings calls, the next step is to link professionals in WRDS Professional to those in Execucomp and to corporate participants in earnings calls.

To match executives from Execucomp to professionals in WRDS Professional, I proceed in three main steps: (i) for each year and company in Execucomp, determine which professionals in WRDS Professional were working at that company during that year; (ii) for each executive in Execucomp, find a match among the professionals in (i) using first name, second, name, middle name, surname, and year born; and (iii) ensure consistency of the database after

[29] Same list as in the previous section.

steps (i) and (ii) are performed for all entries in Execucomp. The specific algorithm used is presented in more detail in Appendix C.4.

To match corporate speakers from Refinitiv to professionals in WRDS Professional, I proceed similarly in three main steps: (i) for each earnings call, determine which professionals in WRDS Professional were working at the company during the call's quarter; (ii) for each corporate speaker in an earnings calls find a match among professionals in (i) using first name, second, name, middle name, and surname; and (iii) ensure consistency of the database after steps (i) and (ii) are performed for all entries in Refinitiv. The specific algorithm used is presented in more detail in Appendix C.5.

### C.4. Linking WRDS Professional and Execucomp

1. Link Execucomp *gvkey* to WRDS *companyid*.
2. For each year $t$ and company $i$ in Execucomp, list executives in that database by their id, *execid*. This is set $Executives_{t,i}$.
3. Obtain the set of professionals in WRDS professional that were working company $i$ in year $t$. This is set $Professionals_{t,i}$.
4. For each executive *execid* in set $Executives_{t,i}$, find the best match in set $Professionals_{t,i}$, using first name, second name, middle name, last name, and year born.
5. When a match is found — say *proid* —, save the executive's id, *execid*, as a new field of information in entry *proid* of WRDS Professional database.
6. After repeating (b)-(e) for every year, company, and executive in Execucomp, it is time to check for inconsistencies in entries. For each *proid* in WRDS Professional, check inconsistencies in matched *execid*. For example: different *execid*'s with different "year born" matched to the same *proid* entry. Because there are very few cases, I perform a manual audit and correction.
7. For each $persoind$ in WRDS Professional, check inconsistencies in matched *execid*. For example: different *execid*'s with different "year born" matched to the same *proid* entry. Because there are very few cases, I perform a manual audit and correction.

8. Import information from Execucomp — such as gender and reason why professional left the company — to WRDS Professional using matched *execid*.

### C.5. Linking WRDS Professional and earnings calls from Refinitiv

1. Keep only corporate speakers.

2. Link Refinitiv *CUSIP* to WRDS *companyid*.

3. For each earnings call in date $t$ for company $i$ in quarterly earnings calls, list corporate speakers by their id, *speakerid*. This is set $CorpSpeakers_{i,t}$.

4. Obtain the set of professionals in WRDS Professional that were working in company $i$ and date $t$. This is set $Professionals_{i,t}$

5. For each speaker *speakerid* in set $CorpSpeakers_{t,i}$, find the best match in set $Professionals_{i,t}$, using first name, middle name, last name.

6. When a match is found — say *proid* —, save the speaker's id, *speakerid*, as a new field of information in entry *proid* of WRDS Professional database.

7. After repeating (c)-(e) repeating for every year, company, and speaker in QEC, it is time to check for inconsistencies in entries. For each *speakerid* in WRDS Professional, check inconsistencies in matched *proid*. For example: same *proid*'s with different "gender" matched to the same *speakerid* entry. Because there are very few cases, I perform a manual audit and correction.

8. For each $persoind$ in WRDS Professional, check inconsistencies in matched *speakerid*. For example: different *execid*'s with different "year born" matched to the same *proid* entry. Because there are very few cases, I perform a manual audit and correction.

9. Import information on gender from QEC to WRDS Professional using matched *speakerid*.

### C.6. Linking IBES to other databases

The next step is linking IBES company identifiers — ticker and CUSIP — to other databases' identifiers: gvkey (Compustat, Execucomp, and Refinitiv earnings calls), permno (CRSP), and companyid (WRDS Professional). This is done by first matching ticker and permno based on CUSIP and company names. Unmatched observations are then matched based on IBES ticker and CRSP exchange ticker, CUSIP, company names, and date. In both, matching of company name is done using Python package "*fuzzywuzzy*" that attributes a positive score (the best score is 0) to string matches based on the Levenshtein distance between them.[30] After linking to permno, we can use links of permno's to gvkeys (based on CRSP-Compustat merged) and to WRDS Professional company id's, described in the next section.

## D. Firm performance in female- and male-led companies

In this appendix section, I investigate whether there are objective differences in firm performance by CEO gender. The key finding is that there are no such differences — at least in the statistical sense —, except for a smaller persistence in one measure of firm performance for female-led companies relative to their male-led peers. Note that while I perform this exercise for the sake of completeness, differences in objective firm performance should not matter for my results. Indeed, my results indicate that analysts' beliefs deviate from the rational expectations benchmark. Hence, as long as analysts correctly perceive any potential differences in objective firm performance between male- and female-led companies, no under- or over-reaction should occur.

To make a comprehensive analysis of potential differences across male- and female-led companies, I measure firm performance in three different ways. I use the earnings per share ratio (normalized by past prices) as described in section 2.2, as well as the year-over-year, $\frac{EPS_t - EPS_{t-4}}{EPS_{t-4}}$, and quarter-over-quarter, $\frac{EPS_t - EPS_{t-1}}{EPS_{t-1}}$, growth rates of the earnings per share.[31]

I also evaluate potential differences in firm performance based on various moments of firm performance: average, volatility, and persistence. With regard to the former, and for

[30] In a nutshell, this distance is determined by how many alterations (additions and deletions) would be necessary to make both strings equal to one another.

[31] Note that I only compute growth rates for the sub-sample of strictly positive earnings per share.

each measure of performance, I run the following regression

$$Y_{i,t} = \theta_m + \theta_f 1(\text{Female-led})_{i,t} + \theta_z Z_{i,t} + u_{i,t}, \tag{18}$$

where $Y_{i,t}$ captures some measure of performance for firm $i$ at quarter $t$, and $Z_{i,t}$ includes firm and quarter-year fixed effects. Regarding the persistence of performance, I extend the regression in (18) to include past performance:

$$Y_{i,t} = \varphi_0 + \varphi_1 1(\text{Female-led})_{i,t} + \varphi_m Y_{i,t-1} + \varphi_f Y_{i,t-1} \times 1(\text{Female-led})_{i,t} + \varphi_z Z_{i,t} + u_{i,t}, \tag{19}$$

To evaluate the volatility of performance, I first compute the sample standard deviation of performance for each CEO-firm pair. Suppose executive $c$ is the CEO of firm $i$ between quarters $t_0$ and $T$, then I compute

$$\text{sd}(Y_{i,c}) = \sqrt{\frac{1}{T - t_0} \sum_{t_0 \leq t \leq T} (Y_{i,t} - \bar{Y}_{i,c})^2}, \text{ where } \bar{Y}_{i,c} = \frac{1}{T - t_0 + 1} \sum_{t_0 \leq t \leq T} Y_{i,t}.$$

Next, I run the following regression:

$$\text{sd}(Y_{i,c}) = \kappa_m + \kappa_f 1(\text{Female-led})_{i,c} + \kappa_z Z_i + u_{i,c}, \tag{20}$$

where $Z_i$ includes firm fixed effects. In particular, I am interested in coefficients $\theta_f$, $\kappa_f$, and $\vartheta_f$ which capture the differences in, respectively, the average, the volatility, and the persistence of firm performance by CEO gender.

Table A3 displays the results. Panel A shows that the average firm performance across female- and male-led companies is not statistically different from one another across the three measures. Similarly, panel B explores the volatility of firm performance with the estimated difference between female- and male-led companies — captured by $\kappa_f$ — not statistically different from zero across the three measures.

Finally, panel C reports estimates for the first-order autocorrelation parameters $\varphi_m$ and $\varphi_f$. The QoQ growth rate for male-led companies displays an autocorrelation of -0.184 ($p < 0.001$), with an estimated difference to female-led companies — captured by $\varphi_f$ — of -0.010 ($p = 0.560$). The estimated autocorrelation for the series of $\frac{EPS}{price}$ is of 0.619 ($p < 0.001$), with an estimated difference of -0.018 ($p = 0.585$) to female-led companies. Regression results for the YoY growth rate are the only to suggest a statistically significant difference of roughly 25% between male- and female-led companies. In particular, the estimated values for $\varphi_m$ and $\varphi_f$ are, respectively, 0.234 ($p < 0.001$), and -0.063 ($p = 0.042$). Note, however, that if we test for differences in performance between male- and female-led companies across these 9 dimensions, using an individual significance level of 5% and applying the Bonferroni adjustment ($5\%/9 \approx 0.6\%$), the hypothesis of no difference between these companies would not be rejected.

Having explored *average* differences in firm performance across male- and female-led companies, I now investigate if a CEO's gender is systematically associated with the state of the world faced by their companies *around the time of their appointment*. To see why this is relevant, imagine that analysts pay little attention to bad news, perhaps because they believe that the probability of a company being in bad states of the world is too low to warrant the cost of paying attention. Now suppose that female CEOs — but not male CEOs — are systematically appointed before bad times of the world. In this case, analysts could find it worthwhile to pay more attention to bad news coming from female-led companies since these female CEOs would signal bad states of the world. This would imply in analysts being more reactive to bad news coming from these female-led companies relative to those of male-led companies.

The testable implication of this scenario is that companies with female CEOs, at least in the first periods after their appointment, should be systematically associated with a performance that is worse than that of their male-led peers.

To test this implication, I run the following regression:

$$
\begin{aligned}
Y_{i,t} =& \vartheta_{cons} + \\
& \sum_{k=-5}^{5} \vartheta_{k,m} \times 1(\text{years since } T = k)_{i,t} + \\
& \sum_{k=-5}^{5} \vartheta_{k,f} \times 1(\text{male-to-female transition at } T)_{i,t} \times 1(\text{years since } T = k)_{i,t} + \\
& + \vartheta_z \times Z_{i,t} + u_{i,t},
\end{aligned} \tag{21}
$$

where $Z_{i,t}$ captures firm fixed effects and quarter fixed effects, and $Y_{i,t}$ is a measure of firm performance. Only male-to-male and male-to-female transitions are included in this regression, which implies that $\vartheta_k$ captures the difference between the average value of $Y_{i,t}$ associated with male-to-female and male-to-male transitions for the $k-$th year since the appointment of a CEO.

The solid blue lines in Figure B7 show estimates for $\vartheta_k$ with their respective 95% confidence interval (dotted lines). If female CEOs are systematically appointed before bad times of the world, we would expect negative and statistically significant estimates for $\vartheta_k$ when $k \in \{1, 2, 3\}$. In other words, we would expect the performance of female-led companies after their appointment to be systematically worse than that of their male-led peers. This is not supported by the data.

The solid blue lines in panels (a), (b) and (c) show that, after the appointment of a new CEO, estimates for $\vartheta_k$ are small and not statistically significant. Moreover, the null hypothesis of whether the average coefficient $\vartheta_k$ before the appointment of a CEO equals the average coefficient after this event is not rejected, with p-values of 0.302, 0.893, and 0.585, respectively, for panels (a), (b), and (c). Note that the interpretation of coefficient $\vartheta_k$ is not causal — rather just correlational — because the appointment of a CEO and the gender of that CEO might be endogenous to firm performance.

## E. Under-estimating the persistence of poor firm performance

I just documented evidence suggestive of the yearly growth of female-led companies being less persistent than that of male-led companies. In this appendix section, motivated by this fact, I explore a simple model where agents under-estimate the persistence of negative yearly growth — poor performance — for male-led companies but not for female-led companies. Such under-estimation could happen if, for example, male CEOs always to choose to disclose the bad news slowly, over the course of many quarters, and analysts don't anticipate this. Meanwhile, female-led companies might "rip off the band-aid" and disclose all the bad news at once.

We can illustrate this scenario with a simple model. Let $s_t$ be the earnings in quarter $t$, which follows the true data generating process

$$s_t = \rho s_{t-1} + \varepsilon_t, \tag{22}$$

where $\varepsilon_t$ is *iid* and normally distributed. Analysts, however, underestimate the persistence of bad earnings results — negative YoY earnings growth. In particular, they form their forecasts according to the following statistical model

$$\begin{aligned} s_{t+1} &= \rho s_t + e_{t+1} \text{ if } s_t \geq s_{t-4}, \\ s_{t+1} &= \rho_b s_t + e_{t+1} \text{ if } s_t < s_{t-4}, \end{aligned} \tag{23}$$

where $\rho_b < \rho$.

Panel (a) of Appendix Figure B6 shows the results from a simulation of this model for the parameter space of $\rho_b$, considering a value of $\rho = 0.745$ (the value estimated in Table A3). This panel shows the mean and respective 5th and 95th percentiles of the distribution of the reaction coefficient of the Coibion-Gorodnichenko regression. It is clear that there is under-reaction for values $\rho_b < \rho$, and over-reaction for $\rho_b > \rho$.

While this is promising to rationalize the results of Section 3.1, there is a main testable implication that is rejected in the data. To understand why, note that this model implies that the under-reaction is concentrated in states with negative yearly earnings growth. panel (b) displays the reaction coefficient from the Coibion-Gorodnichenko for the parameter space of $\rho_b$. We see that for $\rho_b < \rho$, the under-reaction is concentrated in the red line, after negative yearly growth. There is no under- or over-reaction after positive yearly growth.

This testable implication is not supported by the data. Appendix Table A12 shows the results of regressions of forecast errors on forecast revisions for male-led companies by sign of the yearly earnings growth. At both the analyst- and consensus-level, the reaction coefficient after positive yearly growth — which should be zero under the model above — is positive and sizable, having around the same magnitude as the coefficient in baseline results, in Table 2. It is, however, imprecisely estimated, with, for example, p-values of respectively 0.161 and 0.270 in columns 1 (unconditional version) and 2 (conditional on firm, broker, analyst, and forecast period fixed effects). At the analyst level, the difference between reaction coefficients after positive and negative yearly growth is larger than zero — as would be expected under the model — although not statistically significant (with p-values of 0.672 and 0.861, respectively, in columns 1 and 2). However, at the consensus level, these estimates are negative — the opposite direction predicted by the model.

## F. Average forecast errors by CEO gender

In section 3, I showed that there is a difference in analysts' reaction to news about firm performance conditional on CEO gender and that this is a mistake — that is, it represents a deviation from the rational expectations benchmark. In this appendix section, I show evidence suggestive of a second type of mistake: forecast errors do not necessarily average zero conditional on CEO gender. In particular, analysts are more optimistic about male-led companies relative to their female-led peers.

To investigate whether forecast errors differ systematically by CEO gender, I run the following regression:

$$\text{Forecast Error}_{a,i,t+1} = \alpha_m + \alpha_f 1(\text{Female-led})_{i,t} + \alpha_z Z_{a,i,t} + u_{a,i,t}, \tag{24}$$

where $1(\text{Female-led})_{i,t}$ is an indicator function that takes the value of 1 if company $i$ in period $t$ has a female CEO or co-CEO, $Z_{a,i,t}$ includes a series of fixed effects and controls, and $u_{a,i,t}$ are residuals.

Appendix Table A13 suggests that forecast errors are more negative for male-led companies relative to their female-led peers. In other words, it suggests that analysts are, on average, more optimistic about companies led by male CEOs than about those led by female CEOs, as they are on average more negatively surprised by the actual result. Column 2 controls for a firm's size — measured by its end-of-period market capitalization at quarter $t-1$ — and level of earnings — measured by $X_{i,t-1}$. Column 3 further includes four different sets of fixed effects: (i) forecast period $t$ fixed effects (e.g. fixed effects for the 1st quarter of year 2000); (ii) analyst fixed effects; (iii) fixed effects for each company $i$ whose performance is being forecast; and (iv) fixed effects for the broker that each analyst $a$ belongs to. Finally, to understand if these average forecast errors differ depending on whether there was good or bad news, column 4 additionally controls for the sign of the past forecast revision and its interaction with the female-led indicator.

Throughout specifications (columns 1 through 3), the average forecast error for male-led companies ranges from an earnings yield of -0.127 p.p. (p = 0.000) to -0.138 p.p. (p = 0.133), suggesting that analysts are on average optimistic about these male-led companies — that is, their forecasts are on average larger than realizations. In contrast, the average forecast error for female-led companies ranges from an earnings yield of 0.034 p.p. (p = 0.698) to 0.036 p.p. (p = 0.605). The difference between these average forecast errors by CEO gender is statistically significant at least at the 10% level throughout specifications. Interestingly, the results in column 4 suggest that this optimism for male-led companies seems to be concentrated after negative forecast revisions (bad news).

These findings are consistent with the evidence in Jannati, Kumar, Niessen-Ruenzi, and Wolfers (2023) for annual forecasts. The authors document that forecast errors — that is, the difference between the realization and the forecast — are more positive for female-led companies relative to that of their male-led peers. They also find that there is a higher likelihood of analysts recommending to sell the stock of these female-led companies relative to that of their male-led peers.

Based on the simple model introduced in section 3.3, we can rationalize these differences in average forecast errors by assuming that agents may be mistaken about the average firm fundamental value, say $\hat{\mu}_{\theta_{bad}} > \mu_{\theta_{bad}}$ for male-led companies. Introducing such mistake about the mean does not affect the key results shown in that section. Indeed, the Coibion-Gorodnichenko reaction coefficients *conditional on the state of the world* still satisfy equation (9) — that is, depend only on mistakes associated with the standard deviation of the variables in the model. Appendix Figure B5 shows the results from simulations with $\hat{\mu}_{\theta_{bad}} < \mu_{\theta_{bad}}$ while maintaining the parametrization of section 3.3 and setting $\hat{\sigma}_{\theta_{bad}} = 2.5 \times \sigma_{\theta_{bad}}$, with the finding that the mistake about this mean parameter has only small effects on baseline result (when $\hat{\mu}_{\theta_{bad}} = \mu_{\theta_{bad}} = 0$).