

DIAGNOSING DISCRIMINATION: STOCK RETURNS AND CEO GENDER

Justin Wolfers

The Wharton School, University of
Pennsylvania

Abstract

A vast labor literature has found evidence of a “glass ceiling,” whereby women are under-represented among senior management. A key question remains the extent to which this reflects unobserved differences in productivity, preferences, prejudice, or systematically biased beliefs about the ability of female managers. Disentangling these theories would require data on productivity, on the preferences of those who interact with managers, and on perceptions of productivity. Financial markets provide continuous measures of the market’s perception of the value of firms, taking account of the beliefs of market participants about the ability of the men and women in senior management. As such, financial data hold the promise of potentially providing insight into the presence of mistake-based discrimination. Specifically, if female-headed firms were systematically under-estimated, this would suggest that female-headed firms would outperform expectations, yielding excess returns. Examining data on S&P 1500 firms over the period 1992–2004 I find no systematic differences in returns to holding stock in female-headed firms, although this result reflects the weak statistical power of our test, rather than a strong inference that financial markets either do or do not under-estimate female CEOs. (JEL: J7, G14, J44)

1. Introduction

A massive literature examining the labor market progress of women has found evidence of a “glass ceiling,” whereby women are under-represented in the top ranks of management. Moreover, this result is usually robust to the inclusion of a rich set of control variables. Broadly speaking, three sets of explanations have

Acknowledgments: I would like to thank Marianne Bertrand, Alessandra Fogli, Larry Katz, Ulrike Malmendier, Paul Oyer, Joshua Pollet, Stephen Ross, Betsey Stevenson, Geoffrey Tate, Stefano Della Vigna, Grace Wong, and Eric Zitzewitz for their valuable input, and Brett Danaher and Brandon Verblow for outstanding research assistance. Thanks also to seminar audiences at the NBER Labor Studies Summer Institute, Society of Labor Economists, the European Summer Symposium in Labor Economics, and the European Economics Association annual conference. Wolfers is also affiliated with CEPR, IZA, and the NBER. The financial assistance of a Geewax, Terker, and Company Research fellowship and the support of the Zull/Lurie Real Estate Center are gratefully acknowledged.

E-mail address: jwolfers@nber.org

been offered to explain these observations. First, there may be unobserved differences in productivity or preferences that are correlated with gender. The second explanation relates discrimination to taste: Animus by co-workers or customers may be such that the firm's marginal revenue product from promoting women is lower, or alternatively an employer may be willing to accept lower profits in order to avoid promoting women. And third, it may be that the ability of women is systematically mis-assessed. Disentangling these theories would require data on productivity, on the preferences of those who interact with workers, and on perceptions of productivity.

Financial markets provide continuous measures of the market's perception of the value of firms, taking account of the beliefs of market participants about the abilities of men and women in senior management. As such, financial data hold the promise of potentially providing insight into the dynamics of discrimination, and specifically the persistence of biased beliefs about ability.

Thus this paper proposes a test of mistake-based discrimination based on returns in financial markets. Imagine that we ran a market in which traders bet on the performance of both male and female managers. If expectations about the ability of women were systematically biased, then this would suggest that continually betting on strong female performance would yield excess returns. Although explicit prediction markets on the performance of male and female managers do not exist, in equity markets traders do take large positions based at least partly on their assessments of the ability of the CEO. The joint null of efficient markets and no mis-assessment yields a particularly sharp set of testable implications: Risk-adjusted returns to holding stock in female-headed firms should be equal to those of male-headed companies. This paper tests whether financial data suggest that female-headed firms are priced as though the market under-estimates their true worth.

Evidence of mistake-based discrimination is potentially interesting for a variety of reasons. First, mistake-based discrimination is surely inefficient, whereas the case against taste- or statistical-based discrimination requires one to take a stronger ethical position than simply appealing to a utilitarian calculus. Second, evidence that financial markets under-estimate the ability of women undermines the argument that capital markets are sufficiently profit-motivated that costly discrimination will not persist in equilibrium. And third, excess returns can only persist if biased evaluations of ability are ubiquitous. Even if many investors underestimate female managers, excess returns should provide a sufficient inducement for the unbiased minority to buy female-headed firms, undoing the effects of the biased majority. This latter argument is also crucial to thinking about the effects of discrimination in the labor market. If only a few employers are biased, then in equilibrium we will see sorting of women toward unbiased employers; if bias is more ubiquitous then a gender wage gap may emerge.

Beyond the discrimination literature, this inquiry also touches on several related research agendas. Behavioral finance scholars have examined systematic patterns of mistaken inference by investors, but that literature has yet to put much emphasis on inferences made by investors about the company's leadership. Interestingly there is now a large literature in corporate finance emphasizing the importance of CEO characteristics as determinants of firm decisions and hence outcomes. Bertrand and Schoar (2003) examine a range of characteristics, including the age of the CEO and where they obtained their MBA; Bertrand et al. (2004) examine their social and political connections, and Malmendier and Tate (2005) examine their optimism. Each finds large effects of CEO characteristics on firm performance. Other research at the intersection of gender and finance includes: Bertrand and Hallock's (2001) analysis of the relative pay of male and female CEO's which suggests that the wage disparity is largely explained by job and worker characteristics; Lee and James's (2003) event study evidence that markets react negatively to the appointment of female CEOs; Mohan and Chen's (2004) finding that IPO pricing is unrelated to CEO gender, and evidence from Welbourne (1999) and Catalyst (2004) suggesting relatively better financial performance among firms with many senior women.

In ongoing research, Wolfers (2006) examines analysts earnings forecasts, finding that forecast errors are systematically larger for female-headed firms. This bias is also particularly strong among male analysts. The logic of these financial market based tests also extends to other domains, and in ongoing work, I am examining sports betting data to determine whether a similar bias applies against (or for) African American sportsmen.

2. Sample

The most challenging data limitation is finding a sufficiently large (and systematic) collection of data on CEO gender. I focus on CEOs simply because they are the most visible and arguably the most important company employee. The only systematic collection of CEO gender that I am aware of comes from the Execu-comp data, which tracks S&P 1500 firms from 1992–2004. These data identified a total of 64 female CEOs and 4175 male CEOs. Over this 15-year sample 1.3% of CEO-years were worked by women.

Table 1 shows summary statistics by CEO gender, weighting each CEO by their tenure. The samples of male- and female-headed firms are surprisingly similar in terms of size, book-to-market value, price-earnings ratios, firm-level betas, and even industry compositions. Although there are some differences—men tend to head slightly larger and more cyclical firms whereas women are over-represented in retail trade and information sectors—these differences should not be exaggerated given the very small sample of female CEOs.

TABLE 1. CEO gender and firm characteristics—weighted by CEO tenure, 1992–2004.

Firm characteristic	Female-headed firm	Male-headed firm
Market capitalization (\$US, 2000)		
10th percentile	\$112m	\$155m
50th percentile	\$582m	\$967m
90th percentile	\$8,710m	\$9,463m
Employees		
10th percentile	293	445
50th percentile	3,076	4,470
90th percentile	25,000	39,000
Book-to-market value		
10th percentile	0.12	0.14
50th percentile	0.45	0.26
90th percentile	1.06	0.92
Price/Earnings Ratio		
10th percentile	−9.3	−8.6
50th percentile	17.7	16.5
90th percentile	52.2	44.0
Beta (Year-end)		
10th percentile	0.26	0.25
50th percentile	0.81	0.86
90th percentile	1.70	1.71
Industry (NAICS) [Number of CEOs]		
Agriculture	0% [0]	0.4% [16]
Mining	0% [0]	4.2% [153]
Utilities	2.6% [2]	5.7% [227]
Construction	0.7% [1]	1.2% [49]
Manufacturing	41.6% [25]	42.2% [1,808]
Wholesale Trade	0% [0]	3.5% [154]
Retail Trade	17.1% [10]	6.5% [280]
Transport	0% [0]	3.0% [116]
Information	12.4% [10]	8.4% [378]
Finance & Insurance	7.6% [2]	13.1% [491]
Real Estate	1.0% [1]	0.8% [33]
Professional Services	6.3% [7]	3.4% [158]
Administration	6.4% [4]	2.0% [104]
Education	0% [0]	0.3% [11]
Health Care	3.1% [2]	1.9% [83]
Arts & Entertainment	0% [0]	0.4% [19]
Accommodation & Food	0% [0]	2.4% [101]
Other Services	1.1% [1]	0.3% [20]
Unclassified	0% [0]	0.4% [14]
CEO Tenure		
10th percentile	0.6 years	0.6 years
50th percentile	2.8 years	3.2 years
90th percentile	8.5 years	8.4 years
CEO appointed		
10th percentile	Pre-1992	Pre-1992
50th percentile	June 1995	December 1992
90th percentile	July 2001	March 2000

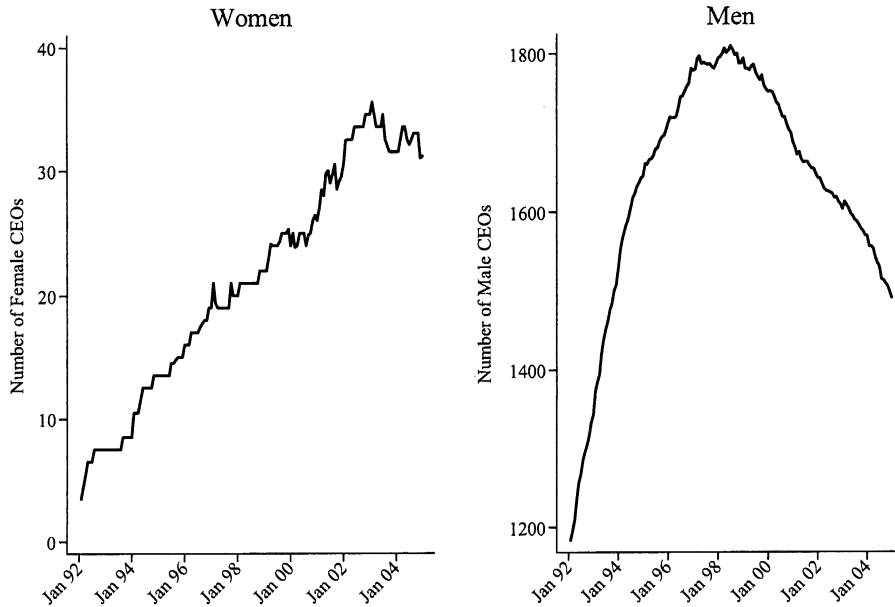


FIGURE 1. CEO gender through time: Execucomp sample.

There are two important respects in which female-headed firms are different from those headed by men. First, as Figure 1 shows, the number of female CEOs has been creeping up through time, from a low of 4 in January 1992, to 34 in December 1994. (The time-series variation in male CEOs in our sample largely reflects the changing coverage of CEO start and end dates in the Execucomp sample.) Female CEOs tend to have had fairly similar tenure, but have been appointed to their positions more recently.

The next two sections examine stock returns data, asking: Do female-headed firms outperform market expectations?

3. Long-term Stock Returns

Several past studies have focused on measuring market expectations regarding the performance of female CEOs. For instance, Lee and James (2003) find that the announcement of a new female CEO leads to a negative abnormal return of 3.7% over the announcement window, whereas a new male CEO leads to only a 0.5% decline in the stock price. Their analysis is potentially informative about market perceptions of the ability of female CEOs, but it does not tell us whether these perceptions turn out to be accurate. As such, the event study evidence is consistent with a view that female CEOs perform as well as men, but that markets systematically underestimate the ability of women. Thus this paper asks instead: Do female-headed firms consistently beat market expectations?

TABLE 2. Performance of the zero-investment portfolio.

Zero Investment Portfolio: Long female-headed firms; Short male-headed firms								
<i>Monthly Returns_t = α + β₁[*](Market_t - R_t^f) + β₂[*]SMB + β₃[*]HML_t + β₄[*]UMD_t + ε_t</i>								
	Value-weighted portfolio				Equal-weighted portfolio			
Alpha	-0.0035 (.0047)	-0.0054 (.0046)	-0.0048 (.0045)	-0.0033 (.0046)	-0.0031 (.0030)	-0.0034 (.0030)	-0.0028 (.0031)	-0.0011 (.0031)
Market-R ^f (VWRF)		0.3065*** (.1097)	0.1416 (.1196)	0.0924 (.1230)		0.0596 (.0710)	-0.0081 (.0833)	-0.0630 (.0846)
Size (SMB)			0.5299*** (.1214)	0.5593*** (.1223)			0.1315 (.0846)	0.1643* (.0841)
Value (HML)			-0.1493 (.1505)	-0.1722 (.1504)			-0.0937 (.1048)	-0.1193 (.1035)
Momentum (UMD)				-0.1347 (.0859)				-0.1502** (.0591)
R ²	.0000	.0483	.2059	.2186	.0000	.0045	.0402	.0796
N	156	156	156	156	156	156	156	156

Notes: (Standard errors reported in parentheses)

*Statistically significant at 10%.

**Statistically significant at 5%.

***Statistically significant at 1%.

Market return measured as excess return of CRSP-weighted index minus the one-month Treasury rate. Size, Value, Momentum are factor returns downloaded from Kenneth French's homepage.

One obvious way to test this would be to assess whether one could earn excess returns by holding stock in female-headed firms relative to holding male-headed firms. Table 2 reports on the returns to holding zero investment portfolios that follow a strategy of buying female-headed firms and shorting male-headed firms. In columns 1–4 we form these portfolios weighting positions proportionately to market capitalization (measured two months prior) within the long or short part of the portfolio; in columns 5–8, portfolio shares are proportionate to the number of female- or male-headed firms in the sample at that point in time. Both of these strategies yield a series of returns which are then regressed on standard factor return series. The remaining differential, α , is the excess return to this portfolio, conditional on risk factors.

Excess returns to following this strategy are typically somewhat negative, with point estimates suggesting negative monthly returns of between one- and three-tenths of a percentage point. Equally, in no case are these returns statistically different from the efficient markets null of zero. Not surprisingly this portfolio is somewhat exposed to the size factor, although momentum also helps explain returns.

In Table 3 I turn to a related approach for measuring excess returns, running a series of Fama-MacBeth regressions. Specifically, for each month in the sample I run a regression:

$$Returns_{i,t} = \alpha_t + \alpha_t^{Female*} Female\ CEO_{i,t} + Z_{i,t}\lambda + \varepsilon_{i,t}$$

where i denotes an individual company, t denotes a month in my 1992–2004 sample, *Return* is the monthly stock return from CRSP, the variable *Female CEO*

TABLE 3. Fama–Macbeth estimates.

	Value-weighted regression			Equal-weighted regression		
Female CEO	-0.0035 (.0047)	0.0023 (.0039)	0.0007 (.0034)	-0.0031 (.0030)	-0.0025 (.0035)	-0.0049 (.0034)
Female CEO & Tenure <4 years		-0.0050 (.0068)	-0.0041 (.0055)		-0.0006 (.0039)	-0.0028 (.0032)
CEO Tenure <4 years		0.0008 (.0030)	0.0031 (.0031)		0.0042 (.0030)	0.0033 (.0028)
Nasdaq firm?			0.0057*** (.0023)			0.0049*** (.0017)
S&P 500 firm?			0.0008 (.0015)			0.0059*** (.0013)
Book-to-Market _{t-1}			0.0045 (.0028)			-0.0002 (.0016)
Ln(Market Cap _{t-2})			-0.0002 (.0018)			-0.0020 (.0016)
Stock Price _{t-2}			0.0004 (.0006)			-0.0041*** (.0018)
Volume _{t-2}			-0.0009 (.0019)			-0.0010 (.0016)
Dividend Yield _{t-1}			0.0469 (.0446)			-0.0209 (.0280)
Returns (t - 2 to t - 3)			0.0004 (.0074)			0.0102** (.0057)
Returns (t - 4 to t - 6)			0.0060 (.0063)			0.0136*** (.0046)
Returns (t - 7 to t - 12)			0.0117*** (.0039)			-.0108*** (.0027)

Notes: (Standard errors reported in parentheses)
 *denote statistically significant at 10%.
 **denote statistically significant at 5%.
 ***denote statistically significant at 1%.

is a variable denoting the proportion of the month in which the firm was run by a women, and $Z_{i,t}$ is a vector of the usual firm characteristics included as controls: Whether the firm is traded on the Nasdaq, whether it is a member of the S&P500, its book-to-market ratio and dividend yield (both measured at the end of the most recent fiscal year), its stock price, the natural logs of market capitalization and trading volume (the latter three variables measured with a two month lag), and three standard momentum variables: Returns over the previous 2–3 months, 4–6 months, and 7–12 months. I follow Gompers, Ishii, and Metrick’s (2003) data appendix in constructing the panel of firm characteristics.

The coefficient α^{Female} is a measure of the differential performance of female-headed firms in that month, controlling for other factors that affected the cross-section of returns. Running these regressions separately for each month yields a time series for α^{Female} , the differential return to holding female-headed firms.

Column one shows the average coefficient in a very sparse regression in which each month’s returns are modeled only as a function only of CEO gender. This result is—by construction—identical to that shown in Table 2. Because we

TABLE 4. Matching estimates—Average monthly returns: female-headed firms and matching male-headed firms.

	Matching by industry then size	Matching by industry, whether book-to-market is above median, and then size
Returns: Female-headed companies	0.0129 (.0028)	0.0132 (.0028)
Returns: Matched male-headed firms	0.0149 (.0027)	0.0150 (.0027)
Difference	-0.0020 (.0034)	-0.0018 (.0034)
Sample size	3263	3261

Note: (Standard error reported in parentheses).

might expect the true ability of a CEO to be revealed during the first few years of their tenure (and hence any mis-assessments to be reversed during that period), the specifications in columns 2 and 5 also include a dummy variable for whether a CEO was in the first four years of his or her tenure, and the same variable, interacted with the female CEO variable. More complete specifications are reported in columns three and six, which not only control for other characteristics, but also allow the effects of these characteristics to vary through time. Whereas the first three columns show the average results through time from running value-weighted cross-sectional regressions each month, the final three columns report the average results from unweighted regressions.

As with Table 2, these results show no statistically significant difference in the returns to holding stock in female-headed firms. Moreover, although one might expect such an effect to be most strongly present in the first few years of a CEO's tenure, there is no statistically significant evidence of this, either.

Although the Fama-Macbeth approach allows one to control for many firm characteristics, it relies on a particular functional form to facilitate regression-adjusted comparisons of all female-headed firms to male-headed firms. It may be more informative to compare the performance of firms that are—at least on observables—otherwise very similar. As such, Table 4 reports on a matching approach. Specifically, for each female-headed firm in each month, I begin by stratifying by the 49-industry groups described in Fama and French (1997) ensuring that I only match female-headed firms with a male-headed firm within the same narrow industry. I then match each female-headed firm to the male-headed firm whose market capitalization two months earlier was most similar. Thus the unit of observation now is not the monthly return to an entire portfolio, but rather a comparison of the return to holding a female-headed firm versus its match partner. This comparison is shown in column 1. In column 2, I sort first by industry, and then within industry sort the firms into whether their book-to-market ratios (measured two months earlier) were above or below the industry median in the sample for that month, and within this double-sort, the matched firm is then

the male-headed firm whose market capitalization (two months prior) was most similar to that of the female-headed firm.

In sum, across the three approaches to measuring differential returns—portfolio returns, Fama-Macbeth regressions, and matching estimates—the difference in returns between female- and male-headed firms are not statistically discernible from zero. As such, these results are consistent with the hypothesis that markets do not systematically under-estimate female-headed firms. However, it is worth emphasizing just how imprecise these results are, as a close inspection of the standard errors reveals that these results do not constitute a particularly convincing demonstration of the absence of discrimination. Under each approach the estimate of the differential monthly returns has a standard error of around one-quarter to one-half of a percentage point, suggesting that even differences in annual returns of as much as 5% would not be detectable. That is, the sample of female CEOs of S&P 1500 firms is just too small to find evidence of excess long-period returns to female-headed firms, even if the misperceptions held by traders are quite large. The next section makes this point more directly.

4. Interpretation: Lack of Power

Although the tests of statistical significance presented above do not reject a null of zero (and hence of no discrimination), this is not the only interesting hypothesis worth considering. Specifically, the sign of the estimated effects is in most cases negative, which may be inconsistent with a view that markets systematically underestimate the ability of female CEOs. We now turn to asking whether these estimates statistically falsify a prior that female CEOs are substantially underestimated. In order to do so, one needs to be more precise about such a prior.

In this example, we will consider the implications of investors who falsely believe that female CEOs produce 10% lower corporate earnings than male CEOs. Traders' valuations will reflect their expectations of the net present value of the firm's earnings. Thus, were a female CEO to live forever, the valuation of her firm would be 10% lower than the value of a male-headed firm. However, the investor takes account of the fact that average CEO tenure is about 10 years, and hence both discounts the future (at 0.5% per month) and each month assigns a 0.8% chance of a change in CEO. Based on CEO hires since 2000, the investor believes that there is about a 2.5% chance that the new CEO will be a woman (whom he would also underestimate), and for simplicity, assumes that this is independent of whether the previous CEO was female. Embedding these assumptions in a monthly model generates a 3.9% discount for female-headed firms. Thus these illustrative numbers suggest that the announcement of a female CEO should lead stock prices to decline by nearly 4%, which is close to Lee and James's (2003) event study estimate. Naturally, the event study tells us only about market perceptions of ability.

Now if the perceptions of lower ability turn out to be false, what are the implications for earnings and capital gains associated with holding female-headed firms? Each quarter the firm will announce earnings 10% higher than the market expected. This is obviously reasonably large relative to earnings surprises, and so examining returns around earnings announcement dates should have some power to find this, a point emphasized in Wolfers (2006).

However the implications for long-term stock returns are less easy to discern and depend on whether investors learn about the CEO's true ability. If investors stick stubbornly to their biased beliefs even in the face of accumulating evidence of good performance by the firm under the female CEO's stewardship, the returns to holding the stock will only be higher by the amount by which earnings in that year exceeded expectations. Given the perception of a 10% ability differential and a 6% annual discount rate (which pins down the price-earnings ratio), this suggests excess annual returns of only 0.6%, although given the persistence of misperceptions, these excess returns are enjoyed every year. Of course, this represents the extreme case where investors do not update their beliefs at all. In the opposite extreme, if investors learn the CEO's true ability in her first period in the job, the initial "female discount" will be reversed immediately, yielding maximum excess returns of 3.9% in the first period, and zero thereafter. It might be more realistic to believe that it takes investors four years to falsify their initial misperceptions about the ability of a specific female CEO. If this is the case, the initial 3.9% decline in value following the appointment of a female CEO would be unwound within four years, yielding around an extra 1% per year capital gain, as well as annual earnings surprises that start at 0.6%, declining to 0% by the fourth year.

All told, this example suggests that even if financial markets substantially under-estimate the ability of female CEOs (in this example by 10%), then excess returns to holding stock in female-headed firms is likely to be less than 1.5 percent per year or around 0.1% per month, and to last for only a few years. As such, even despite the fact that the point estimates in Tables 2–4 suggest negative excess returns to holding stock in female-headed firms, the standard errors are sufficiently large that they cannot reject a null that excess returns are sufficiently positive as to be consistent with substantial bias against female CEOs. Thus, the best conclusion from this analysis is that differences in long-term returns to holding stock in female- and male-headed companies cannot (yet) falsify a null hypothesis that investors substantially under- or overestimate the ability of female CEOs.

This lack of statistical power suggests two alternative research strategies for those interested in estimating whether financial markets underestimate female CEOs. First, we could simply wait for more data to come in. As it stands, the female CEOs in my sample have generated a total of 3,369 female CEO-months; monthly (market-model adjusted) returns have a standard deviation of

around 15%, and as such measuring excess returns in female-headed firms yields a standard error of around $0.15/\sqrt{3168} = 0.26\%$ per month, or 3% per year. In order to get the standard error down to 1%, we would need around ten times this sample size. Obviously, extending the sample back through time will not help much. Going forward, there are about 35 women heading S&P 1500 firms, and this is not increasing particularly rapidly (over the period 2000–2004, 2.5% of new CEO appointments within the S&P 1500 were women, suggesting that until this hiring rate changes, we are currently close to the steady state). As such, it will take until around the year 2020 to generate a sample of 10,000 female CEO-months.

Given these calculations, it seems much more promising to examine whether the quarterly earnings of female-headed firms outperform market expectations, or to examine stock returns specifically around the times that earnings surprises are announced. This is the approach pursued in Wolfers (2006), and the results so far appear reasonably promising.

References

- Bertrand, Marianne, and Kevin F. Hallock (2001). "The Gender Gap in Top Corporate Jobs." *Industrial and Labor Relations Review*, 55(1), 3–21.
- Bertrand, Marianne, Francis Kramarz, Antoinette Schoar, and David Thesmar (2004). "Politically Connected CEOs and Corporate Outcomes: Evidence from France." Working paper, University of Chicago.
- Bertrand, Marianne, and Antoinette Schoar (2003). "Managing with Style: The Effect of Managers on Firm Policies." *Quarterly Journal of Economics*, 118(4), 1169–1208.
- Catalyst (2004). *The Bottom Line: Connecting Corporate Performance and Gender Diversity*. Catalyst, New York.
- Fama, Eugene F., and Kenneth R. French (1997). "Industry Costs of Equity." *Journal of Financial Economics*, 63, 153–194.
- Gompers, Paul, Joy Ishii, and Andrew Metrick (2003). "Corporate Governance and Equity Prices." *Quarterly Journal of Economics*, 118(1), 107–155.
- Lee, Peggy M., and Erika Hayes James (2003). "She'-E-Os: Gender Effects and Stock Price Reactions to the Announcements of Top Executive Appointments." Darden Business School Working Paper No. 02-11.
- Malmendier, Ulrike, and Geoffrey Tate (2005). "CEO Overconfidence and Corporate Investment." *Journal of Finance*, 60(6), 2661–2700.
- Mohan, Nancy, and Carl Chen (2004). "Are IPOs Priced Differently Based on Gender?" *Journal of Behavioral Finance*, 5(1), 57–65.
- Welbourne, Theresa (1999). "Wall Street Likes Its Women: An Examination of Women in the Top Management Teams of Initial Public Offerings." Cornell Center for Advanced Human Resource Studies Working Paper No. 99-07.
- Wolfers, Justin (2006). "Do Financial Markets Discriminate? Evidence from Female CEOs." Working paper, University of Pennsylvania.

