Economics Letters 117 (2012) 436–440

Contents lists available at SciVerse ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Underperformance by female CEOs: A more powerful test

Gueorgui I. Kolev *

EDHEC Business School, 24 avenue Gustave Delory, CS 50411, F-59057 Roubaix Cedex 1, France

ARTICLE INFO

Article history:
Received 29 August 2008
Received in revised form
19 June 2012
Accepted 20 June 2012
Available online 26 June 2012

JEL classification:
G1
G3

Keywords:
Female Chief Executive Officers' underperformance
Linear factor models
Cluster sample econometric methods
Management style

ABSTRACT

Female CEOs underperform their male counterparts in terms of shareholders' returns by roughly 0.35% per month. This difference is significant, comparable to the in-sample value premium, somewhat smaller than the equity and momentum premia, and larger than the size premium.

1. Introduction

In an interesting and provocative article ("As leaders, women rule") reporter Rochelle Sharpe (Businessweek online, 2000) reviews the "findings of a growing number of comprehensive management studies conducted by consultants". The essential outcome is that female executives are rated higher than their male counterparts on various measures, e.g., producing high quality work, goal setting and mentoring employees, regardless of whether the ratings are done by peers, subordinates, or bosses. Ms. Sharpe points out an interesting puzzle: "if women are so great, why aren't more of them running the big companies"?

In this essay I make a small step towards resolving this puzzle by, first, presenting an empirical finding, and second, suggesting possible explanations. The finding is that, *in terms of shareholders' returns*, female Chief Executive Officers (CEOs) underperform their male counterparts by about 0.35% per month. This difference is significant, of magnitude comparable to the in-sample value premium, somewhat smaller than the equity risk premium and the momentum premium, and larger than the size premium. Moreover, the effect is conditional on risk factors that are known to affect stock returns: market risk (measured by the stock market β), and exposures to size, value, and momentum. Possible explanations are suggested in the conclusions.

This study is a close follow-up to a paper by Wolfers (2006), who was the first to document female CEOs long term[1] underperformance of similar magnitude, however statistically *insignificant*. Wolfers (2006) takes as a starting point the fact that firms hiring female CEOs experience more negative price reactions in the event window than firms hiring male CEOs (Lee and James, 2007) and hence conjectures that this is caused by discrimination, i.e., investors (incorrectly) underestimating the managerial skills of female CEOs drive down the stock price in the short term. Hence he conjectures that as investors who are shrewd and free of gender biases correct the hypothesized discrimination driven mispricing, we should observe that female managed firms experience positive abnormal stock returns over the long run. However, finding that the stock of female managed firms underperforms, he focuses on the fact that the effect is insignificant in his regressions and pessimistically concludes that we will need 20 years more of data until we are able to reject the null that females are discriminated by the stock market.

In what follows I show that when information in the sample available is used more efficiently, we can measure female CEOs underperformance quite accurately. In fact, I propose different econometric methods for measuring risk adjusted performance of

* Tel.: +33 0 320 154 558.
E-mail address: gueorgui.kolev@edhec.edu.

[1] By "long term" I mean to distinguish the underperformance documented here and in Wolfers (2006) from results from an event study, the negative price reaction to the announcement of the hiring of a female CEO (Lee and James, 2007). My sample spans about 20 years of monthly data, and in itself, is not particularly long.

0165-1765/$ – see front matter
doi:10.1016/j.econlet.2012.06.028

firms grouped by a certain characteristic (the gender of the CEO in this application). Wolfers (2006) employs calendar time portfolio regressions (also known as Jensen's-α approach, after Jensen, 1968 and Black et al., 1972) and Fama and MacBeth (1973) regressions. The tools he uses are extremely popular in empirical asset pricing, and they are reliable tools which will deliver correctly sized tests in the face of arbitrary cross sectional correlation.

However the calendar time portfolio and Fama–MacBeth regressions in a certain sense weight equally time periods, and switch to using the time period as a unit of observation when estimating precision.[2] This is a reasonable approach when the panel of firms is balanced or when the number of firms in each portfolio is large. However, these methods are likely to result in suboptimal use of the sample information when the panel of firms is strongly unbalanced and the number of firms in one or more of the portfolios is small. A strongly unbalanced panel and few firms in the female portfolio are characteristics of the current application. Very few firms have a female CEO overall and the number of such firms varies a lot as pointed out by Wolfers (2006, Fig. 1). Hence it is important in this application to focus on the return of the given firm in the given month, rather than on the average return of firms in the given month. Intuitively, a month in which there are only 3 firms having female CEOs is less informative about the relative performance of male vs. female CEOs compared to a month in which there are 30 firms with female CEOs. (Loughran and Ritter, 2000 offer a similar critique of the low power of the calendar time portfolio approach with a focus on events that are clustered at certain times, e.g., initial public offerings.)

In the approach that I follow, I use the full information in the cross section of time series, while still computing standard errors correctly in the presence of arbitrary cross sectional correlation (the existence of which is taken as a fact in empirical asset pricing) and heteroskedasticity. This is sufficient to show that the coefficient measuring the differential in long term performance between female and male CEOs is significant.

Next, I employ the Generalized Least Squares (GLS) estimation method that has been advocated by theoretical econometricians (e.g., Wooldridge, 2002, 2006). It amounts to estimating GLS with a weighting matrix taking the random effects (equicorrelated) structure and computing the variance of the estimates that is correct under arbitrary cross sectional correlation and heteroskedasticity. This method is generally consistent under slightly stronger assumptions, which in this context are not binding,[3] and will deliver higher precision if the true and unknown cross sectional correlation structure is well approximated by the random effects structure. This is expected *a priori* to be the case here. Most of the strong and apparently not dying out correlation between returns (even after controlling for what asset pricing models tell us we should control for) probably arises from a common shock shared by all firms in a given period. This shock is not captured by the asset pricing model, e.g., shocks to aggregate risk aversion, changes in social norms and sentiment, neighborhood effects, and herd behavior.

The letter proceeds as follows. In Section 2 I present the econometric model. Section 3 contains the results of the estimation. I postpone the interpretation of the results to the last section.

2. Econometric model

To study how returns of firms led respectively by female and male CEOs differ, we should control for the factors that are known to determine stock returns. In other words, we should compare *risk adjusted* returns. From the Capital Asset Pricing Model (CAPM) we know that a major determinant of the firm specific required return is systematic risk, measured by the stock's market β. Therefore I allow in the regressions below male and female headed firms to have different βs, i.e., different exposures to systematic risk.

Small firms tend to command higher returns (Banz, 1981; Reinganum, 1981). Rosenberg et al. (1985) show that firms with high book-to-market ratios command higher returns. Fama and French (1992, 1993) show that size and book-to-market are responsible for most of the differential return, and after they are accounted for the other anomalies known at the time, e.g., earnings-to-price ratio, sales growth, dividend yield, etc., do not contribute much to explaining stock returns. Jegadeesh and Titman (1993) show that the momentum effect is an important determinant of stock returns.

Following Fama and French (1996) the three factor model has mostly replaced the Capital Asset Pricing Model (CAPM) as an empirical model of expected stock returns. The three factor model includes as risk factors the market return in excess of the risk free rate (MktRf), the size risk factor (SMB), and the value risk factor (HML). Carhart (1997) adds also the momentum factor (MOM) to account for the momentum effect.[4] The four factor model is widely used for performance evaluation and for calculation of risk adjusted returns. The CAPM and the four factor model predict respectively that for any firm or portfolio i, the expected return is given by

$$\mathrm{E}r_i = \alpha + \beta_i\,\mathrm{E(MktRf)} \tag{1}$$

$$\begin{aligned}\mathrm{E}r_i \;=\; &\alpha + \beta_i\,\mathrm{E(MktRf)} + \xi_i\,\mathrm{E(SMB)} + \chi_i\,\mathrm{E(HML)}\\ &+ \mu_i\,\mathrm{E(MOM)},\end{aligned} \tag{2}$$

where E is the expectation operator, and α is the average risk free rate.

To use Eq. (2) for comparing risk adjusted performance of female vs. male CEOs, one can proceed as in Wolfers (2006, Table 2). One constructs the time series of value or equal weighted average returns on a zero investment portfolio which is long in firms having a female CEO and short in firms having a male CEO, and regresses these returns on MktRf, SMB, HML and MOM. Taking the difference between Eq. (2) written once for male and once for female CEOs, we see that the constant in this regression (of the dependent variable which is the zero investment portfolio return long in female and short in male CEO firms, on the independent variables MktRf, SMB, HML and MOM) has to be equal to 0. This constant is known in finance as Jensen's α, and it measures the risk adjusted differential performance between female and male CEOs. Notice that the inference from this regression is robust to arbitrary cross sectional correlation of returns. However this regression switches from the unit of observation being firm/time period to the unit of

2 The framework of the calendar time portfolio approach is known in econometrics as multivariate regression. The details of statistical testing are developed in Gibbons et al. (1989). For example, if a researcher wants to study, as in this application, the performance of firms having female CEOs vs. those having male CEOs, she would start from a panel of firms, and she would form one portfolio of male and one portfolio of female firms. This is accomplished by calculating for each time period the value (or equal) weighted average return by CEO gender. So the researcher reduces the full panel of firm returns to only two time series, one series of average returns for female CEOs, and one series of average returns for male CEOs. Having formed the two time series, the researcher proceeds by estimation and hypothesis testing as in standard multivariate regression analysis.

3 Ordinary Least Squares is consistent under weak exogeneity; basically the error term in the equation for a given executive should be uncorrelated with the regressors of this particular executive. GLS requires for consistency that the error is uncorrelated with the regressors of the given executive and on the top of that uncorrelated with the regressors in the equation of the other executives in the same time period.

4 The risk factors are zero investment portfolios which are, roughly speaking, long in small firms and short in big firms (SMB), long in high book-to-market ratio firms and short in low book-to-market ratio firms (HML), and long in past winners and short in past losers (MOM). Details on the factors' construction and time series of the factors are available from Kenneth R. French's website.

observation being time period, and is blind to the fact that in one period we might have 1 firm led by a female CEO, and in another period we might have 100 firms led by female CEOs.

I propose an alternative approach. For a description of the Generalized Least Squares (GLS) version of the approach consider the following clustered linear model

$$\begin{aligned} r_{ft} = {} & \alpha + \gamma\ 1[\text{female}_{ft}] + \beta_1 \text{MktRf}_t \\ & + \beta_2\ 1[\text{female}_{ft}] * \text{MktRf}_t + \xi_1 \text{SMB}_t \\ & + \xi_2\ 1[\text{female}_{ft}] * \text{SMB}_t + \chi_1 \text{HML}_t \\ & + \chi_2\ 1[\text{female}_{ft}] * \text{HML}_t \\ & + \mu_1 \text{MOM}_t + \mu_2\ 1[\text{female}_{ft}] * \text{MOM}_t + \tau_t + \varepsilon_{ft} \end{aligned} \tag{3}$$

where the subscript f denotes a firm and t denotes a time period (month). The dependent variable r_{ft} is the net return on firm f in period t. The term $1[\text{female}_{ft}]$ is an indicator equal to 1 if the CEO is a female and 0 otherwise, so that the coefficient γ is of the main interest and corresponds to Jensen's α. It measures the differential performance of female CEOs over their male counterparts, upon controlling for the other covariates. MktRf_t is the excess return on the value weighted stock market over the 1 month T-bill, which is changing only across time. The interaction term of the female indicator and the market return captures the different market risk β that the female managed and male managed firms might have, so the β_1 measures the risk of the firms led by male CEOs, and $(\beta_1+\beta_2)$ measures the market risk of the firms headed by female CEOs. The same is true for the other risk factors—we allow male and female headed firms to have different exposure to each of these factors.[5] Note that there are different numbers of firms in each period, so $t = 1, 2, 3, \ldots, T$ and $f = 1, 2, \ldots, F_t$.

If we stack the regressand and regressors in every period, R_t and X_t of dimensionality (F_t by 1) and (F_t by K) respectively, where K is the total number of regressors, and we collect the parameters in a column vector β which is $K \times 1$, the model can be written as

$$R_t = X_t\beta + e_t, \tag{4}$$

e_t being the composite error and $e_{ft} \equiv \tau_t + \varepsilon_{ft}$. I assume strict exogeneity

$$\text{E}[e_{ft}|X_t] = 0$$

basically requiring that the error for any CEO is mean independent of the regressors for that particular CEO and additionally of the regressors of any other CEO in period t, which seems plausible. Then the random effects estimator b of the parameter vector β will be consistent:

$$b = \left[\Sigma_{t=1,\ldots,T}(X_t'S_t^{-1}X_t)\right]^{-1}\left[\Sigma_{t=1,\ldots,T}(X_t'S_t^{-1}R_t)\right], \tag{5}$$

$S_t \equiv s_\tau^2 1_{Ft}' 1_{Ft} + s_\epsilon^2 I_{Ft}$, where 1 is a (row) vector of ones and I is an identity matrix, both with dimensionality given in the subscript. The matrix S is a consistent estimator of the random effects weighting matrix $\text{Var}\ e_t = \sigma_\tau^2 1_{Ft}' 1_{Ft} + \sigma_\epsilon^2 I_{Ft}$.

The part of the error term τ_t has the standard for random effects model properties—independent across time and common for all firms in period t. The idiosyncratic error ε_{ft} is *not* assumed to have the standard properties—it is allowed to be arbitrarily correlated across firms within the time period, and heteroskedastic. The only assumption made, which is typical of asset pricing applications is that ε_{ft} (and hence e_t) is independent across time. To conduct valid inference under this set of assumptions, I need the estimated variance matrix of b fully robust to heteroskedasticity and arbitrary within cluster correlation (here a cluster is the time period t):

$$\begin{aligned} \widehat{\text{Avar}}\ b = {} & \left[\Sigma_{t=1,\ldots,T}(X_t'S_t^{-1}X_t)\right]^{-1}\left[\Sigma_{t=1,\ldots,T}(X_t'S_t^{-1}\hat{e}_t\hat{e}_t'S_t^{-1}X_t)\right] \\ & \times \left[\Sigma_{t=1,\ldots,T}(X_t'S_t^{-1}X_t)\right]^{-1}, \end{aligned} \tag{6}$$

where $\hat{e}_t = R_t - X_t b$ is the GLS residual.[6]

The description of the Ordinary Least Squares (OLS) version of the approach is similar to the above, except that the time specific random effect τ_t is omitted in Eq. (3), and the identity matrix I_t of appropriate dimension replaces S_t in Eqs. (5) and (6).

3. Results

I merge data from Execucomp on the gender of the CEO and when he/she began and ended his/her service as a company CEO, CRSP data on stock returns, and time series of the risk factors MktRf, SMB, HML and MOM from Kenneth R. French's website.[7] For summary statistics see Table 1.

The dependent variable is the firm specific monthly total shareholders' return including dividend distributions in percentage form. The main regressor of interest is the Female CEO indicator variable which reflects the differential performance of the female CEOs relative to the base group of male CEOs. Note that the excess return on the value weighted market portfolio over the risk free rate is included plus its interaction with the Female CEO indicator—this controls for the possibility that female headed firms have different market β from the male headed firms. Similarly I include both main effects for SMB, HML, MOM, and their interaction with the Female CEO indicator variable (see Table 2).

Unconditionally, female CEOs generate lower returns, but the effect is not significant, column (1) and column (2). Conditioning on the market factor, we obtain somewhat mixed results. In column (3) estimated by OLS, the differential performance of female CEOs is insignificant, i.e., the coefficient on the Female indicator is insignificant; however firms with female CEOs appear to be more risky, as reflected by market β of 0.11 higher than the market β of male CEOs. In column (4) estimated by GLS the result is reversed—female CEOs underperform and the effect is marginally significant; however female CEOs do not appear to have higher market β.

From column (5) to column (10) female CEOs underperformance controlling for risk factors is significant, and ranges in magnitude from 0.28% to 0.41% per month. For a sense of whether this is large or small we can compare to the average in-sample premia on the risk factors MktRf, SMB, HML, and MOM in Table 1. Apparently female CEOs underperformance is an important discount, of magnitude larger than the small firm premium, similar to the value premium, and somewhat smaller than the equity and momentum premia.

Firms with female CEOs do not have different exposure to the HML factor compared to firms with male CEOs. Also the difference in market β between female and male CEOs becomes small and insignificant in the specifications from column (5) to column (10).

Firms with female CEOs do have more extreme exposures to the SMB and MOM factors compared to firms with male CEOs. Apparently firms having female CEOs behave like small firms (higher exposure to SMB than firms with male CEOs) that are past losers (lower exposure to MOM than firms with male CEOs).

4. Discussion and conclusion

Based on my sample, the fact is that female CEOs underperform their male counterparts. What does this mean? I suggest several possible explanations.

5 As the anonymous referee pointed out, this is analogous to the Oaxaca–Blinder decomposition familiar from the labor economics literature.

6 The generic elements of each weighting matrix $S_t,\ t = 1, \ldots, T$ are the same. But the dimensionality of the matrix differs across t, as the sample is unbalanced and there are different numbers of firms in each time period.

7 All the data were downloaded on 7th of May, 2012.

Table 1
Summary statistics.

Variable	Mean	Std. dev.	Min.	Max.	Obs.
Return (total shareholders' return in percent)	1.55	14.36	−98.13	937.36	491 375
Female (female CEO indicator)	0.01	0.12	0	1	491 375
MktRf (market premium in percent)	0.49	4.51	−18.55	11.53	240
SMB (size premium in percent)	0.24	3.52	−16.62	22.06	240
HML (value premium in percent)	0.36	3.38	−12.87	13.88	240
MOM (momentum premium in percent)	0.54	5.29	−34.75	18.4	240

Table 2
The dependent variable is Return, the monthly total shareholders' return including dividends, in percentage form (mean = 1.55%; standard deviation = 14.36%).

	(1) OLS	(2) GLS	(3) OLS	(4) GLS	(5) OLS	(6) GLS	(7) OLS	(8) GLS	(9) OLS	(10) GLS
Female	−0.40	−0.27	−0.28	−0.29*	−0.41**	−0.34**	−0.35**	−0.34**	−0.35**	−0.28*
	[0.33]	[0.17]	[0.19]	[0.17]	[0.18]	[0.16]	[0.17]	[0.17]	[0.16]	[0.16]
MktRf			1.12**	1.13**	1.04**	1.05**	1.10**	1.10**	1.04**	1.04**
			[0.04]	[0.04]	[0.04]	[0.04]	[0.03]	[0.03]	[0.02]	[0.02]
Female*MktRf			0.11**	0.06	0.07	0.02	0.01	0.02	−0.06	−0.04
			[0.05]	[0.04]	[0.04]	[0.04]	[0.04]	[0.04]	[0.04]	[0.04]
SMB					0.39**	0.40**	0.51**	0.51**	0.54**	0.54**
					[0.08]	[0.08]	[0.05]	[0.05]	[0.04]	[0.04]
Female*SMB					0.22**	0.19**	0.19**	0.18**	0.22**	0.21**
					[0.07]	[0.06]	[0.06]	[0.06]	[0.06]	[0.06]
HML							0.38**	0.37**	0.33**	0.33**
							[0.04]	[0.04]	[0.03]	[0.03]
Female*HML							−0.01	−0.01	−0.05	−0.05
							[0.06]	[0.05]	[0.05]	[0.06]
MOM									−0.16**	−0.16**
									[0.02]	[0.02]
Female*MOM									−0.13**	−0.13**
									[0.03]	[0.03]
Constant	1.56**	1.53**	0.98**	0.97**	0.93**	0.91**	0.74**	0.73**	0.88**	0.86**
	[0.35]	[0.36]	[0.15]	[0.14]	[0.12]	[0.11]	[0.08]	[0.08]	[0.07]	[0.07]
Observations	491 375	491 375	491 375	491 375	491 375	491 375	491 375	491 375	491 375	491 375

Robust standard errors clustered by month are in brackets. The number of clusters (months) in the sample is 240.
* $p < 0.10$.
** $p < 0.05$.

First, there is a possible circularity in our results in that the measure used to assess CEO performance – i.e., shareholders' returns – is not independent of shareholder beliefs. Thus, if shareholders believe that females are less effective CEOs (Lee and James, 2007), there could be a self-fulfilling prophecy.

A second and related explanation notes that feedback regarding CEO skill is extremely noisy such that almost any belief about a CEO may be maintained for extended periods of time. Thus, if investors start off with the prior that females are worse CEOs than males (however defined), the lack of clear feedback can sustain this belief for long periods of time, and beyond the 20 years my sample covers (Einhorn and Hogarth, 1978).

Third, the regression results reported can be interpreted as suggesting that boards of directors practice affirmative action and pursue socially desirable goals at the expense of their shareholders. It is, of course, still an open question as to whether this is true.

Fourth, firms with female CEOs might be perceived as less risky by investors, and therefore might require lower expected returns to persuade investors to hold equity in a firm having a female CEO. Of course this dimension of male CEO risk premium has to be different from the dimensions of risk that we already control for, i.e., market risk, small firm risk, value risk and momentum risk.

In conclusion, there is a puzzle and our data and analysis do highlight different dimensions of a possible solution. At the same time, they also suggest the need for auxiliary measures of the attitudes of boards of directors and shareholders toward female CEOs.

Acknowledgments

I would like to thank Robin M. Hogarth and the anonymous referee for their comments and suggestions on an earlier draft of this paper. The views expressed and any errors are my own.

References

Banz, R.W., 1981. The relationship between market value and return of common stocks. Journal of Financial Economics 9, 3–18.
Black, F., Jensen, M.C., Scholes, M., 1972. The capital asset pricing model: some empirical tests. In: Jensen, Michael C. (Ed.), Studies in the Theory of Capital Markets. Praeger Publishers Inc.
Carhart, M.M., 1997. On persistence in mutual fund performance. Journal of Finance 52 (1), 57–82.
Einhorn, H.J., Hogarth, R.M., 1978. Confidence in judgment: persistence of the illusion of validity. Psychological Review 85 (5), 395–416.
Fama, E.F., French, K.R., 1992. The cross section of expected returns. Journal of Finance 47, 427–465.
Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fama, E.F., French, K.R., 1996. Multifactor explanations of asset pricing anomalies. Journal of Finance 51 (1), 55–84.
Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: empirical tests. Journal of Political Economy 71 (3), 607–636.
Gibbons, M.R., Ross, S.A., Shanken, J., 1989. A test of the efficiency of a given portfolio. Econometrica 57 (5), 1121–1152.
Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. Journal of Finance 48, 65–91.
Jensen, M.C., 1968. The performance of mutual funds in the period 1945–1964. Journal of Finance 23 (2), 389–416.
Lee, P.M., James, E.H., 2007. She'-*E*-Os: gender effects and investor reactions to the announcements of top executive appointments. Strategic Management Journal 28 (3), 227–241.
Loughran, T., Ritter, J.R., 2000. Uniformly least powerful tests of market efficiency. Journal of Financial Economics 55 (3), 361–389.
Reinganum, M.R., 1981. Misspecification of capital asset pricing: empirical anomalies based on earnings' yields and market values. Journal of Financial Economics 9, 19–46.
Rosenberg, B., Reid, K., Lanstein, R., 1985. Persuasive evidence of market inefficiency. Journal of Portfolio Management 11, 9–17.
Wolfers, J., 2006. Diagnosing discrimination: stock returns and CEO gender. Journal of the European Economic Association 4, 531–541.
Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press.
Wooldridge, J.M., 2006. Cluster sample methods in applied econometrics: an extended analysis.