Journal of Modern Applied Statistical Methods

Volume 7 | Issue 2 Article 21

11-1-2008

Data Mining CEO Compensation

Susan M. Adams
Bentley University, sadams@bentley.edu

Atul Gupta
Bentley University, agupta@bentley.edu

Dominique M. Haughton
Bentley University, dhaughton@bentley.edu

John D. Leeth
Bentley University, JLeeth@Bentley.edu

DOI: 10.22237/jmasm/1225513200

Journal of Modern Applied Statistical Methods
November, 2008, Vol. 7, No. 2, 561-576

1538 – 9472/08/$95.00

Data Mining CEO Compensation

Susan M. Adams Atul Gupta Dominique M. Haughton John D.Leeth
Bentley University

The need to pre-specify expected interactions between variables is an issue in multiple regression. Theoretical and practical considerations make it impossible to pre-specify all possible interactions. The functional form of the dependent variable on the predictors is unknown in many cases. Two ways are described in which the data mining technique Multivariate Adaptive Regression Splines (MARS) can be utilized: first, to obtain possible improvements in model specification, and second, to test for the robustness of findings from a regression analysis. An empirical illustration is provided to show how MARS can be used for both purposes.

Key words: data mining, interactions, modeling, multivariate adaptive regression splines (MARS), multiple regression

Introduction

The use of multiple regression analysis is widespread in empirical research. To use multiple regression analysis the full set of independent variables affecting the dependent variable must first be identified and all of the expected interactions among these explanatory variables specified. Since both theoretical and practical considerations make it impossible to pre-specify all possible interactions, the explanatory power of any given regression specification will be limited. In addition, while theory may provide guidance as to which predictors to use in a model, the functional form of the dependent variable on the predictors is unknown in many cases. This article describes two ways in which the data mining technique Multivariate Adaptive Regression Splines (MARS) can be utilized: first, to obtain possible improvements in model specification, and second, to test for the robustness of findings from a regression analysis. An empirical illustration of how MARS can be used for both purposes is then provided.

The intuition underlying MARS is straightforward; the algorithm examines the data for all possible interactions among the specified explanatory variables and for non-linear relations between the dependent and explanatory variables and, in general, yields substantial improvements in explanatory power. Findings from the MARS analysis can be used in two possible ways. First, MARS may yield insight into possible empirical relationships that exist in data, but which have not been identified by the researcher. Such relationships can be examined for theoretical content and used to improve the specification of the regression model.

A second useful application of MARS is in the context of testing for the robustness of findings from a particular regression. For example, consider a research study interested in examining the relationship between employee gender and compensation. Because compensation is expected to depend on a variety of characteristics, the typical regression model includes a set of explanatory variables and a dummy variable to capture the gender effect.

Susan M. Adams is a Professor in the Department of Management. Email: sadams@bentley.edu. Atul Gupta is a Professor in the Department of Finance. Email: agupta@bentley.edu. Dominique M. Haughton is a Professor in the Department of Mathematical Sciences. Email: dhaughton@bentley.edu. John D. Leeth is a Professor in the Department of Economics. Email: jleeth@bentley.edu.

The sign and statistical significance of the dummy variable and the explanatory power of the entire model depend on three factors: the choice of explanatory variables, the set of interactions included in the model, and the specified functional form of the dependent variable in terms of the predictors. While MARS can add no insight into the choice of explanatory variables, it can test for all possible interactions among the explanatory variables, the preponderance of which have not been included in a normal regression analysis.

Moreover, MARS uses splines (understood here to be piecewise-linear functions) to allow for possible non-linearities in the data. Given that MARS will generally yield a substantial improvement in explanatory power, a finding that the sign and statistical significance of a variable of interest (the dummy variable for employee gender in our example) remains unchanged serves as a useful test for the robustness of the findings from the original regression. Normally, researchers using regression analysis provide the results from several model specifications to demonstrate the empirical strength of their conclusions. MARS provides a more structured approach to this model specification procedure and, thereby, generates a more powerful test of robustness.

Methodology

The Data

Standard and Poor's (S&P) ExecuComp database was used to examine the compensation of male and female CEOs. This database tracks a variety of corporate data for the 1500 largest companies in the U.S. from 1992 to 2003 and personal and compensation data for their associated CEOs. From 1992 through 2003, 56 women served as CEOs of the top 1500 Standard & Poor's companies in the United States; in contrast, 4,242 men served as corporate CEOs over the same time period. The ExecuComp database yielded 214 individual executive/year observations for female CEOs and 18,179 observations for male CEOs. The CEOs are scattered across 369 4-digit SIC industries. To control for possible industry effects in salary determination, analysis focused on CEOs employed in the forty-one 4-digit SIC industries with at least one female CEO.

Table 1 gives a summary of the variables used in the analysis. The left-hand side of the table provides information on the OLS sample and the right-hand side provides information on the MARS sample. To be included in an OLS regression an observation must have a complete set of information on all explanatory variables. The MARS sample is larger because the MARS procedure explicitly controls for missing values, allowing all observations with information on total compensation to be included in the analysis, an important advantage of MARS over OLS.

The dependent variable used was the logarithm of the CEO's total compensation for the year, which includes salary, bonus, restricted stock, stock options (evaluated using the Black-Scholes procedure), long-term incentive payouts, and other types of compensation. The independent variables are fairly standard. Most studies of wages and salaries include information on human capital such as education, general labor market experience, and experience within a specific company (Topel, 1991; Willis, 1986). The ExecuComp data does not provide information on education and measures of experience are somewhat spotty. To capture human capital characteristics included in the analysis are age and the number of years the person has served as CEO. (For some CEOs, the data lists the date the person started working for the company. Unfortunately, the information was available for only 59.2 percent of the sample and so was not used in the analysis.)

Because economic theory indicates that investments in human capital should have positive but diminishing returns, also included were squares of age and years as CEO. While early studies of the pay-performance relationship found little evidence of such a link (see Jensen & Murphy, 1990), some recent work documents that CEO compensation is related to company size and company performance (see Bebchuk & Grinstein, 2005). Company size is measured using the dollar value of sales revenue and company performance using the return on assets.

Table 1: Means (Standard Deviations)

Variable	OLS Sample Men	OLS Sample Women	Difference in Means (absolute t/z statistic)	MARS Sample Men	MARS Sample Women	Difference in Means (absolute t/z statistic)
Total Compensation (thousands of 2003 $)	5,036 (17,332)	4,926 (9,402)	110 (0.15)	4,797 (16,589)	4,768 (9,257)	28 (0.04)
Log Total Compensation	7.69 (1.176)	7.73 (1.151)	-0.04 (0.43)	7.64 (1.170)	7.68 (1.157)	-0.04 (0.45)
Age	53.96 (7.396)	51.14 (7.396)	2.82 (5.30)**	54.13 (7.937)	51.03 (7.327)	3.09 (5.96)**
Years as CEO	8.11 (7.957)	7.97 (11.991)	0.14 (0.16)	8.06 (7.894)	7.93 (11.974)	0.13 (0.15)
Sales (billions of 2003 $)	2.93 (7.228)	2.70 (8.576)	0.23 (0.37)	2.85 (7.033)	2.63 (8.440)	0.22 (0.37)
Return on assets (Percent)	0.10 (29.129)	1.68 (15.773)	-1.58 (1.31)	0.44 (28.033)	1.82 (15.564)	-1.38 (1.19)
Manufacturing	0.406	0.405	0.001 (0.03)	0.413	0.423	-0.010 (0.77)
Transportation	0.149	0.049	0.100 (3.98)**	0.152	0.047	0.105 (4.23)**
Trade	0.072	0.195	-0.123 (6.35)**	0.070	0.188	-0.117 (6.29)**
Finance	0.049	0.078	-0.029 (1.88)	0.051	0.075	-0.024 (1.56)
Services	0.324	0.273	0.051 (1.53)	0.314	0.268	0.047 (1.43)
Number	3,689	205		4,058	213	

Finally, to control for differences in pay across industries and over time the OLS analysis includes binary variables measuring the company's 1-digit SIC code and a linear time trend. The MARS analysis permits a more detailed investigation of industry and time effects. The MARS procedure includes a categorical variable representing 41 different 4-digit SIC industries and a categorical variable representing 12 different years. All dollar figures for total compensation and sales revenue have been adjusted to correct for the impact of inflation and are stated in 2003 dollars.

Table 1 uncovers only a few statistically significant differences in means or proportions between male and female CEOs. Within the four-digit SIC industries examined, female CEOs are a few years younger than their male counterparts and the companies they operate are more likely to be involved in trade and less likely to be involved in transportation. In terms of compensation, the data provide no evidence that male and female CEOs are paid differently.

The MARS methodology

The MARS algorithm, proposed by Friedman in 1991, relies on the following basic ideas:

For each continuous independent variable, MARS creates a piecewise linear function with too many change points (knots) to begin with, and then prunes unnecessary knots by a backward procedure. Consider the functions BF3 and BF4 (Basis Functions 3 and 4) identified by MARS (definitions of all Basis Functions are given in Appendix A). These two functions are preceded by BF1, as follows:

BF1 = (SALES > .);
BF3 = max(0, SALES – 1.747087) * BF1;
BF4 = max(0, 1.747087 - SALES) * BF1;

BF1 is zero whenever the variable SALES is missing, and one otherwise. The functions BF3 and BF 4, taken together, define a piecewise linear function of SALES, with a break point (otherwise referred to as a knot or a change point) at about 1.75 billion dollars. Note that BF3 is zero when SALES is less than 1.747, and BF4 is zero when SALES is greater than 1.747. Basis functions are chosen by MARS to achieve the best fit in a regression of the dependent variable on the Basis Functions. Of course, without any restriction on over-fitting, better and better fits will be attained by using more and more Basis Functions breaking at more and more knots. MARS uses a backward stepwise method to eliminate Basis Functions and knots which contribute least to the fit of the model.

For each independent categorical variable, MARS groups categories and creates dummy variables which correspond to these groups in such a way as to yield the best fit possible. For instance, the Basis Function BF5, given by the expression is:

BF5 = (SICNEW = 1 OR SICNEW = 2 OR SICNEW = 5 OR SICNEW = 13 OR SICNEW = 15 OR SICNEW = 16 OR SICNEW = 21 OR SICNEW = 22 OR SICNEW = 23 OR SICNEW = 25 OR SICNEW = 26 OR SICNEW = 27 OR SICNEW = 28 OR SICNEW = 29 OR SICNEW = 31 OR SICNEW = 32) * BF1;

BF5 equals one if the SICNEW code for an observation is one of those listed in the expression (1, 2, 5, 13, …, etc.), zero otherwise. This means that, of all the ways MARS considered to create a dummy variable that would represent a group of industries, the grouping in BF5 is one of the groupings it found would yield the best fit with the dependent variable. Other industry groupings are identified and expressed in other Basis Functions.

MARS looks for interactions among independent variables, by introducing into the model the product of two variables, if such an interaction leads to a sufficient improvement in the model. For example, the Basis Functions BF23 and BF24 represent an interaction of age with the number of years as CEO since BF21 includes BF18 in its expression, which in turns includes age. An interesting aspect is that MARS can (and often will) create interactions, not between original variables, but between restrictions of these variables to a particular range as is done in BF23 and BF24. BF23 (respectively BF24) interacts age with number of years as CEO, but only beyond 12 years as CEO (respectively up to 12 years as CEO), and in any case only up to ages of 43 years. BF23 and BF24 have a different coefficient in the final model, so the strength of the interaction depends on the range of years as CEO involved in the interaction: it is stronger (.030) for BF24 than for BF23 (.019).

To summarize, MARS ends up with a collection of Basis Functions, which are transformations of independent variables taking into account non-linearities and interactions. MARS then estimates a least-squares model with a parsimonious set of Basis Functions as independent variables. Parsimony is achieved by removing Basis Functions, knots and interactions which do not contribute sufficiently to the model fit.

MARS, in essence, is an OLS procedure, but with judicious transformations of the independent variables. Risks of overfitting are controlled in various ways by the algorithm (Friedman, 1991, Section 3.6). To take into account the fact the data are used not only to estimate the coefficients of the Basis Functions but to create these Basis Functions in the first

place, a penalized sum of squared residuals is minimized to select the final model (in least squares regression, a non-penalized sum of squares would be used). This is achieved by minimizing a quantity referred to as the Generalized Cross Validation (GCV) criterion equal to $(1/N)\ SSR/[1-C(M)/N]^2$ (see Friedman, 1991, p. 20), where N is the number of observations, SSR is the residual sum of squares, and $C(M)$ is a measure of the complexity of a model with M Basis Functions. The complexity $C(M)$, which would equal M in usual least squares modeling, is defined to be equal to $M + dM$, where d is a penalty for each additional Basis Function.

The parameter D can be determined in a number of ways: a value of 3 has been recommended on the basis of simulations in Friedman (1991), but a larger value may be appropriate for larger sample sizes. An alternative, used in this article, is to determine the parameter d via ten-fold cross validation (not to be confused with the GCV mentioned above, the GCV does not actually involve cross-validation). Ten-fold cross- validation involves randomly dividing the data into ten parts, building the model – with various values of the parameter d – with nine tenths of the data, and evaluating the performance of the model on the remaining tenth. This is done ten times, for each tenth in turns, and the performance averaged out over the ten runs. The value of d yielding the best performance is selected, and the GCV criterion is computed with this value of d. A clearly over-fitting model is first built, and Basis Functions are removed one after the other, yielding a sequence of models with a decreasing number of Basis Functions. A model is selected from that sequence which minimizes the GCV criterion.

A convenient place to get information with introductions to the MARS methodology, white papers, and useful references is the Salford Systems Web site (www.salford-systems.com). The article by De Veaux, et al. (1993) includes a good introduction to MARS, albeit in the context of chemical engineering, and contrasts the MARS methodology with that of neural networks. The article by Sephton (2001) gives an introduction to MARS and evaluates how well MARS performs at forecasting recessions; the author finds that for the time series considered for predicting recessions, MARS yields a better in-sample, but a worse out-of-sample performance than for instance probit regression (with a dependent variable of 1 if a time period was in recession, 0 if not); this may indicate that the MARS models used in this context were over-fitting the data to some extent. This is the reason why it is recommended in the literature (Deichman, et. al. 2002; Munoz & Felicisimo 2004) to evaluate MARS on validation samples, independent of the sample used to build the data, in order to select a MARS model that will not over-fit the data and will predict well on validation samples.

This approach is adopted in Deichman, et al. (2002) where MARS is used in the context of direct response modeling; the authors find that response models which use MARS Basis Functions perform better than alternatives on independent validation samples. Munoz & Felicisimo contrast a MARS methodology with several alternatives and reach two interesting conclusions: one is that MARS yields the best predictive power, and the other is that an independent validation sample is truly needed (cross-validation is not sufficient).

The issue of over-fitting is considered later in this article and will explain why in our case over-fitting does not risk calling results into question. Finally, an article where MARS is used in analyses of living standards in Vietnam (see for example Deichman, et. al. (2001)), where interesting interactions are revealed between regions of the country and other predictors when modeling the logarithm of household expenditure per capita, indicating that such models of household wealth are likely to differ across regions, with the importance of some predictors varying across these regions.

Results

Table 2 presents the OLS results. As is typical, several specifications to check for robustness are included. The first specification includes only human capital characteristics, while the second augments these characteristics with information on the company. The third specification controls for differences in pay by industry and over time and the fourth specification interacts each

independent variable with the binary variable indicating the gender of the CEO. The last specification is a test to determine if any significant differences exist in how male and female CEOs are paid across the variables considered. In standard parlance, it is a test to determine if it is permissible to pool male and female CEOs in the same sample.

The results in Table 2 appear remarkably robust. In none of the first three specifications is the female binary variable statistically significant, indicating no difference in pay between male and female CEOs. Although in the fourth specification the F-statistic indicates male and female CEOs are paid differently, the only statistically significant difference in CEO pay is in the transportation industry, but the positive interaction term points to female CEOs earning more than their male counterparts. In short, in terms of pay the data provide no evidence of discrimination against women once they have made it to the highest rung of the corporate ladder. Almost all other studies of gender differences in compensation find women earning far less than men, controlling for other factors including occupation and title (Bertrand & Hallock, 2001).

The other variables in Table 2 are also robust across the four empirical specifications in terms of statistical significance and absolute size. In all four specifications general experience as measured by age raises log total compensation but at a decreasing rate (the coefficient on age is significantly positive and the coefficient on age squared is significantly negative). Company size as measured by sales and company performance as measured by return on assets significantly boost CEO compensation. The positive coefficient on time demonstrates a substantial yearly increase in real CEO compensation and the negative coefficients on transportation and trade shows CEOs in these industries earn less, all else equal, than CEOs in manufacturing (the excluded category). The other variables are insignificant across all four specifications.

Appendix A presents the full set of MARS results. The MARS model explains about 46 percent of the variability in (logged) total compensation, compared to about 17 percent for the OLS model. This improvement is due (in part) to the fact that MARS identifies groups of industries for which the compensation model differs, a matter very much at the heart of compensation modeling, and successfully includes interactions of these industry groupings and other independent variables.

Most important to our analysis, gender does not enter the model at all once the above mentioned interactions are included. Even following a very structured approach for determining model specification, an approach which investigates hundreds of possible interactions among the independent variables and allows for complex non-linear relationships to exist between the dependent and independent variables, the data still uncovers no difference in how male and female CEOs are compensated. A maximum of 80 basis functions were allowed to be used in this MARS model, and ten-fold cross-validation were used to evaluate models considered by MARS. The maximum number of basis functions allowed (80) is sufficient for MARS to build a large enough model from which to prune to get a satisfactory final model (such a maximum should be at least as large as about twice the number of basis functions in the final model; in this case the final model contains 33 basis functions, so an initial maximum of 80 basis functions is ample). To determine how much to prune (in other words how many basis functions to drop) to yield a final model, MARS uses as a measure of performance a modified R-square measure referred to as the Generalized Cross Validation (GCV) criterion; the GCV incorporates a cost per basis function into its formula; the higher the cost, the smaller the number of basis functions in the final model. The choice of that cost is quite crucial, and is performed here by ten-fold cross validation, which consists in splitting the data into ten parts, using 9/10 of the data to build the model and the remaining tenth to evaluate candidate models corresponding to different choices of cost in order to select the cost that yields the best performance on the held out tenth of the data. Typically, and here as well, each tenth of the data plays the role of a hold-out sample in turns and performance is judged on all ten such samples. The absence of a gender effect in CEO

Table 2: OLS Results on Log Total Compensation ($2003)

	(1)	(2)	(3)	(4)
Constant	3.118	4.442	3.806	3.773
	(2.31)*	(3.33)**	(2.91)**	(2.79)**
Female	0.041	0.035	-0.060	2.193
	(0.24)	(0.22)	(0.35)	(0.36)
Age	0.166	0.117	0.133	0.134
	(3.29)**	(2.38)*	(2.71)**	(2.65)**
Age squared	-0.001	-0.001	-0.001	-0.001
	(3.10)**	(2.36)*	(2.63)**	(2.55)*
Years CEO	-0.011	-0.004	-0.011	-0.015
	(0.96)	(0.33)	(1.03)	(1.31)
Years CEO squared	0.000	0.000	0.000	-0.000
	(0.77)	(0.45)	(0.89)	(1.05)
Sales (billions 2003$)		0.049	0.049	0.049
		(5.94)**	(5.76)**	(5.35)**
Return on assets		0.003	0.004	0.004
		(2.80)**	(3.30)**	(3.30)**
Time			0.048	0.050
			(6.13)**	(6.24)**
Transportation			-0.669	-0.694
			(7.92)**	(8.12)**
Trade			-0.289	-0.330
			(2.24)*	(2.37)*
Finance			-0.030	0.007
			(0.19)	(0.04)
Service			-0.070	-0.083
			(0.84)	(0.98)
Age×Female				-0.86
				(0.36)
Age squared×Female				0.001
				(0.30)
Years CEO×Female				0.104
				(1.95)
Years CEO squared×Female				-0.002
				(1.90)
Sales×Female				-0.002
				(0.13)
Return on assets×Female				-0.005
				(0.96)
Time×Female				-0.034
				(0.80)
Transportation×Female				0.967
				(2.72)**
Trade×Female				0.504
				(1.31)
Finance×Female				-0.586
				(1.23)
Service×Female				0.264
				(0.69)

Table 2: OLS Results on Log Total Compensation ($2003) (continued)

	(1)	(2)	(3)	(4)
R-squared	0.02	0.12	0.17	0.18
F-statistic: all coefficients = 0	3.04**	8.59**	14.44**	10.50**
F-statistic: female and female interaction terms = 0				2.93**

* significant at 5%; ** significant at 1% *Note*: The t statistics are calculated using standard errors that correct for heteroskedasticity and the correlation among observations for the same individual. Industry results are measured relative to the excluded category, manufacturing.

compensation is robust across empirical specifications.

Tables 3 and 4 summarize the MARS results. To simplify matters, the two tables present results only for observations in the data set where none of the independent variables are missing. When one or more independent variables is missing, the model adjusts for that in the equations (see for example BF1 in Appendix A, which captures the fact that the variable SALES is not missing), but the adjustments involve a fairly small number of observations (see Table 1).

An examination of the basis functions in Appendix A reveals that, for observations without missing values, MARS identifies fourteen groups of Standard Industry Codes (SIC) among which it determines that the models for (log of) total compensation differ. Table 3 categorizes each of the 41 4-digit SIC industries by MARS-created SIC group. The first column of the table lists the industry's 1-digit SIC code, the second column provides a description of the 4-digit SIC industry, and the final columns of the table identify which of the 14 broadly related MARS industries each 4-digit SIC industry belongs. The effects of the various industry variables on total compensation depend on these industry groups; as seen in Appendix A that a 4-digit industry can appear in multiple MARS groupings since different industry groupings can interact with different independent variables.

Generally, researchers investigating industry effects classify firms based on the firm's 1-digit or 2-digit SIC code. The OLS analysis in Table 2 allows CEO compensation to shift upward or downward depending on the firm's 1-digit SIC industry. The Swiss-cheese appearance of Table 3 indicates, at least in terms of CEO compensation, that industry effects are far more complex than a simple upward or downward shift in compensation. Multiple industry interactions exist among the independent variables and the interactions are not grouped according to 1-digit or 2-digit SIC industry.

Table 4 presents the impact of each of the independent variables by industry. The notation with a plus sign (+) as a superscript indicates the expression in brackets is evaluated only for observations where the expression is positive. The expression is set equal to zero for all other observations. Blanks in the table indicate that the coefficient of the expression in the 1st column is zero for that particular industry group. For example, Panel A demonstrates that, as estimated in the MARS model, in SIC1 a one percentage point increase in a company's return on assets (ROA) raises total CEO compensation by 1.4 percent (0.014 log points) when ROA is below 7.047 percent but by 3.6 percent (0.035 log points) when ROA is above 7.047 percent. (In a log-linear specification a one-unit change in an independent variable causes a $e^{\hat{\beta}} - 1$ percentage change in the dependent variable, where $\hat{\beta}$ is the estimated parameter. For small values, β is approximately equal to the percentage change.) In the second SIC group a one percentage point increase in ROA has no impact on log total CEO compensation when ROA is below 1.206 percent but, surprisingly, reduces total CEO compensation by 8.0 percent (0.077 log points) when ROA is above 1.206 percent. MARS uncovers no significant impact on CEO compensation from higher ROAs in the other 12 industry groups. The OLS regressions presented in Table 2 model pay for performance

as a general phenomenon across industries. The MARS methodology, in contrast, discovers ROA affecting CEO pay in only a few 4-digit SIC industries, meaning that pay for performance is far more limited than one might have originally thought.

The second panel in Table 4 reveals that in all industry groups except for SIC5 and to some extent SIC3, CEO compensation rose over time. The coefficient on year is generally zero from 1992 to 1997 but positive for years 1998 to 2003. The parameter of 0.206 on the years 1998 to 2003 indicates that, all else equal, CEOs earned about 23 percent more in these years than in the years from 1992 to 1997 in industry groups other than SIC3, SIC4, and SIC5. The largest jump in salaries over time occurs in SIC4 where the impact of year moves from a -0.439 log points for years 1992 to 1997 to a +0.206 log points for years 1998 to 2003. Other studies also find a rise in CEO salaries in the 1990s (Bebchuk & Grinstein, 2005). The MARS results indicate not a general upward trend in CEO compensation in the 1990s, as implied by the OLS results in Table 2, but a structural break in compensation occurring in 1998.

As can be seen in Panel C, the impact of an additional year of CEO experience (YRSCEO) depends on the age of the CEO, a rough proxy for general labor market experience, and the overall level of CEO experience. For CEOs younger than 43 an additional year of CEO experience lowers total compensation for individuals serving as CEO for less than 12 years but raises it for individuals serving as CEO for more than 12 years. For CEOs older than 43 an additional year of CEO experience has no impact on total compensation except in SIC2 where the impact of greater CEO experience is positive and SIC3 where the impact of greater CEO experience is negative for individuals serving as CEO for less than 1.63 years.

The MARS results on CEO experience are in contrast to the OLS results in Table 2. OLS finds no impact of CEO experience on total compensation, while MARS discovers additional CEO experience raising compensation in some cases but lowering it in others. The counterintuitive results of CEO experience reducing compensation apply to very few observations in the sample. Only 249 of the sample observations are for CEOs younger than 43 with less than 12 years of CEO experience (6.4 percent) and only 340 observations are for CEOs with less than 1.63 years CEO experience in industry group SIC3 (8.7 percent).

The positive impact of CEO experience on compensation pertains to many more observations: 926 observations in SIC2 have more than 0.583 years of CEO experience and are older than 43 (23.8 percent) and 30 observations are for CEOs younger than 43 with more than 12 years CEO experience (0.8%). For the remaining 2,163 observations (55.5 percent) MARS finds no impact on compensation from greater CEO experience. In other words, the MARS results indicate for the vast majority of CEOs greater CEO experience has either a positive or a neutral impact on compensation although for a few CEOs in some industries and at some levels of general and CEO-specific experience greater years heading the company reduces compensation.

Panel D shows the impact on CEO compensation from increases in company size as measured by sales revenue. As can be seen, the impact of company size depends on the company's current level of sales, the age of the CEO, and the industry. Ignoring the age effect, an increase in sales has a larger impact when a company is small, sales less than $1.7471 billion (70.9 percent of the sample), than when it is large, sales greater than $1.7471.

Age augments the impact of sales on CEO compensation for CEOs older than 43 in companies with less than $8.1352 billion in sales revenue and for CEOs younger than 43 in companies with less than $4.4857 billion in sales revenue. Evaluated at the mean age of 53.8, a $1 billion dollar increase in sales revenue raises CEO compensation in most industry groups by 75.5 percent for companies with sales of less than $1.7471 billion, by 6.3 percent for companies with sales between $1.7471 billion and $8.1352 billion, and by 1.82 percent for companies with sales greater than $8.1352 billion. Mathematically, company size appears to raise CEO compensation but at a decreasing rate.

Table 3. MARS identified industry groups

1-digit SIC Industry	4-digit SIC Industry	SIC1 BF5	SIC2 BF25	SIC3 BF6	SIC4 BF13	SIC5 BF61	SIC6 BF73	SIC7 BF45	SIC8 BF43	SIC9 BF11	SIC10 BF7	SIC11 BF57	SIC12 BF19	SIC13 BF51	SIC14 BF75
Mfg	Broadwoven Fabric Mills, Cotton	■					■			■	■	■			
Mfg	Apparel & Other Finished Prods of Fabrics & Similar Mat'l	■										■			■
Mfg	Men's & Boys' Furnishings, Work Clothing, & Allied Garments		■	■		■	■	■	■	■	■	■	■		
Mfg	Newspapers: Publishing or Publishing & Printing			■	■		■	■			■			■	■
Mfg	Commercial Printing	■					■					■			■
Mfg	Pharmaceutical Preparations			■	■			■	■					■	
Mfg	Biological Products, (No Diagnostic Substances)		■	■	■	■		■				■		■	■
Mfg	Perfumes, Cosmetics & Other Toilet Preparations		■	■	■			■				■			
Mfg	Pottery & Related Products		■	■			■	■	■				■		
Mfg	Special Industry Machinery, NEC			■				■	■		■				
Mfg	Computer & Office Equipment			■	■		■	■			■				
Mfg	Computer Peripheral Equipment, NEC			■		■	■	■	■		■				
Mfg	Electric Housewares & Fans	■	■		■		■		■		■				
Mfg	Telephone & Telegraph Apparatus			■					■				■	■	
Mfg	Motor Vehicle Parts & Accessories	■							■			■			
Mfg	Motor Homes	■							■		■				■
Mfg	Electromedical & Electrotherapeutic Apparatus			■	■		■		■			■			
Mfg	Dolls & Stuffed Toys			■			■	■	■						■
Mfg	Miscellaneous Manufacturing Industries			■		■	■	■	■			■		■	
Trans, Comm & Utilities	Communications Services, NEC			■		■	■		■		■		■		
Trans, Comm & Utilities	Electric Services	■			■				■		■			■	
Trans, Comm & Utilities	Natural Gas Distribution	■			■	■	■			■	■	■			■
Trade	Retail-Apparel & Accessory Stores	■				■		■				■		■	■
Trade	Retail-Women's Clothing Stores			■	■		■	■			■	■			
Trade	Retail-Furniture Stores	■				■				■	■				
Trade	Retail-Drug Stores & Proprietary Stores	■										■			
Trade	Retail-Jewelry Stores	■				■	■			■					
Trade	Retail-Catalog & Mail-Order Houses	■	■							■	■	■			
Finance, Ins, Real Estate	Savings Institution, Federally Chartered	■										■			■
Finance, Ins, Real Estate	Patent Owners & Lessors		■	■					■						■
Services	Services-Personal Services	■				■	■		■			■			■
Services	Services-Help Supply Services	■	■				■						■		■
Services	Services-Computer Programming, Data Processing, etc.			■	■				■				■		
Services	Services-Prepackaged Software		■	■		■			■	■			■		■
Services	Services-Computer Integrated Systems Design			■	■		■	■	■		■				
Services	Services-Telephone Interconnect Systems		■	■											
Services	Services-Business Services, NEC		■	■	■		■	■				■	■	■	
Services	Services-Medical Laboratories		■	■	■			■	■		■	■			■
Services	Services-Child Day Care Services			■				■			■	■			
Services	Services-Research, Accounting, Engineering, Management			■					■				■		■
Services	Services-Commercial Physical & Biological Research			■					■		■	■			■

Table 4: MARS Results on Log Total Compensation

Panel A: Return on Assets (ROA)

	SIC1 BF5	SIC2 BF25	SIC3 BF6	SIC4 BF13	SIC5 BF61	SIC6 BF73	SIC7 BF45	SIC8 BF43	SIC9 BF11	SIC10 BF7	SIC11 BF57	SIC12 BF19	SIC13 BF51	SIC14 BF75
$(ROA-7.047)^+$	0.035													
$(7.047-ROA)^+$	-0.014													
$(ROA-1.206)^+$		-0.077												

Panel B: Year

	SIC1 BF5	SIC2 BF25	SIC3 BF6	SIC4 BF13	SIC5 BF61	SIC6 BF73	SIC7 BF45	SIC8 BF43	SIC9 BF11	SIC10 BF7	SIC11 BF57	SIC12 BF19	SIC13 BF51	SIC14 BF75
Yrs 98-03	0.206	0.206	0.206	0.206	0.206	0.051	0.206	0.206	0.206	0.206	0.206	0.206	0.206	0.206
Yrs 92-97				-0.439	0.225									
Yrs 92,93,95,98,03			-0.228											

Panel C: CEO Tenure (YRSCEO)

	SIC1 BF5	SIC2 BF25	SIC3 BF6	SIC4 BF13	SIC5 BF61	SIC6 BF73	SIC7 BF45	SIC8 BF43	SIC9 BF11	SIC10 BF7	SIC11 BF57	SIC12 BF19	SIC13 BF51	SIC14 BF75
$(YRSCEO-.583)^+$		0.018												
$(1.63-YRSCEO)^+$			0.350											
$(YRSCEO-12.0)^+ x (43-AGE)^+$	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019
$(12-YRSCEO)^+ x (43-AGE)^+$	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030

Panel D: Sales Revenue (SALES), in billions 2003 $

	SIC1 BF5	SIC2 BF25	SIC3 BF6	SIC4 BF13	SIC5 BF61	SIC6 BF73	SIC7 BF45	SIC8 BF43	SIC9 BF11	SIC10 BF7	SIC11 BF57	SIC12 BF19	SIC13 BF51	SIC14 BF75
$(SALES-1.7471)^+$	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	-0.019	0.018	0.018	0.018	0.018	0.077
$(1.7471-SALES)^+$	-0.519	-0.519	-0.519	-0.519	-0.519	-0.519	-0.519	-0.821	-0.519	-0.519	-0.519	-0.519	-0.519	-0.519
$(0.2881-SALES)^+$		-3.000												
$(8.1352-SALES)^+ x(AGE-43)^+$	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004
$(4.4857-SALES)^+ x(43-AGE)^+$	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059

Panel E: Age of the CEO (AGE)

	SIC1 BF5	SIC2 BF25	SIC3 BF6	SIC4 BF13	SIC5 BF61	SIC6 BF73	SIC7 BF45	SIC8 BF43	SIC9 BF11	SIC10 BF7	SIC11 BF57	SIC12 BF19	SIC13 BF51	SIC14 BF75
$(AGE-43)^+$	0.035	0.035	0.035	0.035	0.035	0.035	0.017	0.035	0.035	0.035	0.035	-0.016	0.035	0.035
$(43-AGE)^+$													0.201	
$(AGE-54)^+$	-0.052	-0.052	-0.052	-0.052	-0.052	-0.052	-0.052	-0.052	-0.052	-0.052	-0.052	-0.052	-0.052	-0.052
$(AGE-43)^+ x (8.1352-SALES)^+$	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004
$(43-AGE)^+ x (4.4857-SALES)^+$	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059
$(43-AGE)^+ x (YRSCEO-12)^+$	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019
$(43-AGE)^+ x (12-YRSCEO)^+$	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030

Note: Table 3 lists the specific 4-digit SIC industries comprising each SIC industry grouping. A superscript on a bracketed term indicates the expression is evaluated only for observations where the expression is positive. The expression equals zero for all other observations. Blanks in the table indicate the associated industry effect is zero. The table presents results only for observations with information on all independent variables. Appendix A presents the full set of MARS results including the impact of missing values.

The OLS results in Table 2 examine the impact of sales revenue on CEO compensation but the impact of sales is assumed to be linear. The MARS results suggest that a more appropriate specification would include sales revenue and sales revenue squared to allow for the positive but diminishing returns from company size. (When both sales revenue and sales revenue squared are included in the OLS regression both coefficients are highly statistically significant (p values < 0.001) but the inclusion alters the size and significance of the other coefficients only slightly.) In Table 2 across all specifications, an additional $1 billion of sales revenue creates a 5.0 percent increase in CEO compensation. In Table 3, an additional $1 billion of sales revenue creates in most industries a 6.3 percent increase in CEO compensation when evaluated at the sample means of age and sales revenue ($2.914 billion).

The final panel of Table 4 reports the impact of age on CEO compensation. The last four rows of the Panel E simply duplicate the interactive results on age and sales and age and years as CEO discussed previously. Across all age groups higher sales revenue either expands the positive impact of age on total compensation or contracts the negative impact – the interaction between age and sales is positive. Surprisingly, for CEOs younger than 43 an additional year of general experience as measured by age reduces total compensation, all else equal. The reduction is smaller as years as CEO expands for CEOs serving for fewer than 12 years but is larger as years as CEO expands for CEOs serving more than 12 years. In all but SIC13 an additional year of general experience raises total compensation by 1.42 percent for CEOs from 43 to 54 but reduces total compensation by 7.56 percent for CEOs older than 54 when evaluated at the mean level of sales. The influence of age on total compensation is not impacted by years as CEO for CEOs older than 43.

The stereotypical age/earnings profile has a worker's earnings rising steeply early in his or her career, leveling off over time, and then declining. Researchers include age and age squared as independent variables in an OLS analysis of earnings to capture the positive but diminishing impact of general experience on earnings and to allow for the possibility of earnings hitting a peak at some point. Based on the OLS results, CEO compensation hits a peak somewhere between 54.9 and 57.2 years of age depending on the empirical specification. Although the MARS results do not reproduce the standard leveling off of earnings, they do indicate an earnings peak at age 54, a result largely consistent with the OLS analysis.

Conclusion

In most empirical investigations theory guides the selection of independent variables but rarely dictates the functional relationship between the dependent and the independent variables or specifies all possible interactions among the independent variables. Consequently, researchers generally present several sets of results generated using slightly different estimating relationships to demonstrate that the conclusions of the analysis are robust to model specification. Multivariate Adaptive Regression Splines (MARS) is a data mining technique that examines data for all possible interactions among specified explanatory variables and for non-linear relations between the dependent and explanatory variables. By using MARS researchers can check for the robustness of their empirical findings in a highly structured manner, thereby providing a more convincing case that the results are insensitive to model specification. Additionally, MARS may uncover relationships that can be examined for theoretical content and aid future research in the area.

As an example of how MARS can be used as a procedure to check for robustness and as an aid in future research, we examine data on CEO compensation to determine if pay differences exist between men and women. Most studies find men out earn women by a sizable margin even after controlling for differences in education, experience, and occupation (Altonji & Blank, 1999; Bertrand & Hallock, 2001; Stanley & Jarrell, 1998). Using standard OLS analysis we find no evidence male CEOs have an advantage over female CEOs in terms of compensation. Across the four empirical specifications we examine female CEOs earn the same or more than male CEOs, all else equal. In

the MARS methodology the variable representing gender never enters the model indicating that no significant pay difference exists between male and female CEOs. The MARS model controls for observable characteristics and considers all possible interactions among the observable characteristics and total compensation in addition to potential nonlinearities in the relationships between the observable characteristics and total compensation. In short, the absence of a gender effect on CEO compensation is robust.

In terms of the other factors affecting CEO compensation, OLS generates a fairly standard picture of CEO compensation. All else equal, CEOs leading larger companies as measured by sales revenue, more profitable companies as measured by return on assets, and who have more general labor market experience as measured by age earn more than CEOs leading smaller companies, less profitable companies, and who have less general labor market experience. Over time CEO compensation has expanded by almost 5 percent per year in real terms and CEOs in transportation and trade earn less than CEOs in manufacturing. Inconsistent with the human capital model of earnings, OLS finds no reward for CEO experience.

The MARS results are generally consistent with the OLS results but with some important distinctions. Similar to OLS, MARS finds sizable differences in CEO compensation across industries. Unlike OLS, the MARS grouping of industries is unrelated to a broader industry classification such as a 1- or 2-digit SIC code. Further, the MARS industry effects do not simply increase or decrease compensation but instead interact with the other independent variables, suggesting the underlying model of compensation varies by industry groupings. However, note that these industry groupings are not the recognized industry groups based on 1- or 2-digit SIC codes. Similar to OLS, MARS shows CEO compensation rising over time, but unlike OLS the rise is not gradual. In most of the MARS industry groups a structural break in compensation occurs in 1998 causing CEO pay to jump by about 23 percent. In the OLS analysis, the impact of return on assets is modeled as a general phenomenon across industries. The MARS analysis finds return on assets raising CEO compensation but in only one broad industry grouping – meaning pay for performance is fairly limited. The OLS analysis uncovers a positive, linear relationship between sales revenue and CEO compensation. The MARS results suggest sales revenue has a positive but diminishing impact on CEO compensation. In the OLS analysis, the number of years a person has served as CEO appears to have no impact on compensation, while MARS finds CEO experience raising total compensation but only in a few industry groupings. Finally, OLS indicates a CEO's age, a proxy for general labor market experience, raises total compensation but at a decreasing rate, a result in line with the human capital model and the stereotypical age/earnings profile. MARS finds a far more complex relationship with compensation falling, rising, and then falling again as the CEO ages. Both the OLS and the MARS results imply CEO compensation peaks at around 54 years of age.

It is not suggested that MARS be used as a replacement to the standard procedures of model building and hypothesis testing. Instead, MARS may be viewed as a complement to the more traditional methods of analysis. There are implications for practicing managers to consider when evaluating the use of MARS and OLS. For the manager who wants to understand the dynamics of executive compensation, the MARS model provides more details about the specifics related his or her particular situation (e.g., the industry grouping formed by MARS and corresponding interactions). By examining data for unanticipated and possibly complex interactions among the independent variables and for potential nonlinear relationships between the dependent and independent variables, MARS allows researchers to conduct a structured test of robustness and determine important areas for future research. In particular, the MARS analysis of CEO compensation suggests additional work is required to determine the factors causing the compensation explosion in 1998, the reasons for the paucity of pay for performance, and the elements generating common compensation practices across industries.

References

Altonji, J. G., & Blank, R. M. (1999). Race and gender in the labor market. In O. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics*, *3c*, Amsterdam: Elsevier Science B. V., 3143-3259.

Bebchuk, L. A., & Grinstein, Y. (2005). The growth of executive pay. *Oxford Review of Economic Policy*, *21*, 283-303.

Bertrand, M., & Hallock, K.F. (2001). The gender gap in top corporate jobs. *Industrial and Labor Relations Review*, *55*, 3-21.

De Veaux, R. D., Psichogios, D. C. & Ungar, L. H. (1993). A comparison of two non-parametric schemes: MARS and neural networks. *Computers in Chemical Engineering, 17*, 819-837.

Deichman, J., Haughton, D., Phong, N. & Tung, P.D. (2001). A graphical and statistical analysis of the correlates of poverty in Vietnam in 1993 and 1998. In *Living standards during an economic boom in Vietnam 1993-1998*, UNDP and GSO, Hanoi. Available at http://www.undp.org.vn/undp/docs/2001/living/lse.pdf, accessed January 3rd 2006.

Deichman, J., Eshghi, A., Sayek, S. & Teebagy, N. (2002). Application of multiple adaptive regression splines (MARS) in direct response modeling. *Journal of Interactive Marketing, 16*, 15-27.

Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. *Annals of Statistics, 19*, 1-141.

Jensen, M. C., & Murphy, K. J. (1990). Performance, pay, and top management incentives. *Journal of Political Economy 98*, 225-264.

Munoz, J., & Felicisimo, A. (2004). Comparison of statistical methods commonly used in predictive modeling, *Journal of Vegetation Science, 15*, 285-292.

Salford Systems (2006). http://www.salford-systems.com/mars.php, accessed January 2nd 2006.

Sephton, P. (2001). Forecasting recessions: can we do better on MARS? *Federal Reserve Bank of St. Louis Review, 83*, 39-49.

Stanley, T. D., & Jarrell, S. B. (1998). Gender wage discrimination bias? A meta-regression analysis, *Journal of Human Resources, 33*(4), 947-973.

Willis, R. J. (1986), Wage determinants: A survey and reinterpretation of human capital earnings functions. In O. C. Ashenfelter & R. Layard (Eds.), *Handbook of labor economics*, *1*, Amsterdam: Elsevier Science B.V., 525-602.

Appendix A: The MARS model; basis functions and estimated equation

Basis Functions

```
BF1 = (SALES > .);
BF3 = max(0, SALES – 1.747087) * BF1;
BF4 = max(0, 1.747087 - SALES ) * BF1;
BF5 = (SICNEW = 1 OR SICNEW = 2 OR
      SICNEW = 5 OR SICNEW = 13 OR
      SICNEW = 15 OR SICNEW = 16 OR
      SICNEW = 21 OR SICNEW = 22 OR
      SICNEW = 23 OR SICNEW = 25 OR
      SICNEW = 26 OR SICNEW = 27 OR
      SICNEW = 28 OR SICNEW = 29 OR
      SICNEW = 31 OR SICNEW = 32) * BF1;
BF6 = (SICNEW = 3 OR SICNEW = 4 OR
      SICNEW = 6 OR SICNEW = 7 OR
      SICNEW = 8 OR SICNEW = 9 OR
      SICNEW = 10 OR SICNEW = 11 OR
      SICNEW = 12 OR SICNEW = 14 OR
      SICNEW = 17 OR SICNEW = 18 OR
      SICNEW = 19 OR SICNEW = 20 OR
      SICNEW = 24 OR SICNEW = 30 OR
      SICNEW = 33 OR SICNEW = 34 OR
      SICNEW = 35 OR SICNEW = 36 OR
      SICNEW = 37 OR SICNEW = 38 OR
      SICNEW = 39 OR SICNEW = 40 OR
      SICNEW = 41) * BF1;
BF7 = (SICNEW = 1 OR SICNEW = 3 OR
      SICNEW = 4 OR SICNEW = 10 OR
      SICNEW = 11 OR SICNEW = 12 OR
      SICNEW = 13 OR SICNEW = 16 OR
      SICNEW = 20 OR SICNEW = 21 OR
      SICNEW = 22 OR SICNEW = 24 OR
      SICNEW = 25 OR SICNEW = 28 OR
      SICNEW = 35 OR SICNEW = 38 OR
      SICNEW = 39 OR SICNEW = 41);
BF9 = (YEAR = 1998 OR YEAR = 1999 OR
      YEAR = 2000 OR YEAR = 2001 OR
      YEAR = 2002 OR YEAR = 2003) * BF1;
BF10 = (YEAR = 1992 OR YEAR = 1993 OR
      YEAR = 1994 OR YEAR = 1995 OR
      YEAR = 1996 OR YEAR = 1997) * BF1;
BF11 = (SICNEW = 1 OR SICNEW = 3 OR
      SICNEW = 22 OR SICNEW = 25 OR
      SICNEW = 27 OR SICNEW = 28 OR
      SICNEW = 34) * BF3;
BF13 = (SICNEW = 4 OR SICNEW = 6 OR
      SICNEW = 7 OR SICNEW = 8 OR
      SICNEW = 11 OR SICNEW = 13 OR
      SICNEW = 17 OR SICNEW = 21 OR
      SICNEW = 22 OR SICNEW = 24 OR
      SICNEW = 33 OR SICNEW = 35 OR
      SICNEW = 37 OR SICNEW = 38) *
      BF10;
BF15 = (AGE > .) * BF1;
BF16 = (AGE = .) * BF1;
BF17 = max(0, AGE - 43.000) * BF15;
BF18 = max(0, 43.000 - AGE ) * BF15;
BF19 = (SICNEW = 3 OR SICNEW = 9 OR
      SICNEW = 14 OR SICNEW = 20 OR
      SICNEW = 32 OR SICNEW = 33 OR
      SICNEW = 34 OR SICNEW = 37 OR
      SICNEW = 40) * BF17;
BF21 = (YRSCEO > .) * BF18;
BF23 = max(0, YRSCEO - 11.997) * BF21;
BF24 = max(0, 11.997 - YRSCEO ) * BF21;
BF25 = (SICNEW = 3 OR SICNEW = 7 OR
      SICNEW = 8 OR SICNEW = 9 OR
      SICNEW = 13 OR SICNEW = 28 OR
      SICNEW = 30 OR SICNEW = 32 OR
      SICNEW = 34 OR SICNEW = 36 OR
      SICNEW = 37 OR SICNEW = 38);
BF27 = (SALES > .) * BF25;
BF30 = max(0, 0.288101 - SALES ) * BF27;
BF32 = max(0, 8.135196 - SALES ) * BF17;
BF33 = (YEAR = 1992 OR YEAR = 1993 OR
      YEAR = 1995 OR YEAR = 1998 OR
      YEAR = 2003) * BF6;
BF35 = (ROA > .) * BF5;
BF37 = max(0, ROA - 7.047) * BF35;
BF38 = max(0, 7.047 - ROA ) * BF35;
BF39 = (YRSCEO > .) * BF25;
BF40 = (YRSCEO = .) * BF25;
BF41 = max(0, YRSCEO - 0.583) * BF39;
BF43 = (SICNEW = 3 OR SICNEW = 6 OR
      SICNEW = 9 OR SICNEW = 10 OR
      SICNEW = 12 OR SICNEW = 13 OR
      SICNEW = 14 OR SICNEW = 15 OR
      SICNEW = 16 OR SICNEW = 17 OR
      SICNEW = 18 OR SICNEW = 19 OR
      SICNEW = 20 OR SICNEW = 21 OR
      SICNEW = 30 OR SICNEW = 31 OR
      SICNEW = 33 OR SICNEW = 34 OR
      SICNEW = 35 OR SICNEW = 37 OR
      SICNEW = 38 OR SICNEW = 40 OR
      SICNEW = 41) * BF4;
BF45 = (SICNEW = 3 OR SICNEW = 4 OR
      SICNEW = 6 OR SICNEW = 7 OR
      SICNEW = 8 OR SICNEW = 9 OR
      SICNEW = 10 OR SICNEW = 11 OR
      SICNEW = 12 OR SICNEW = 18 OR
```

SICNEW = 19 OR SICNEW = 23 OR SICNEW = 24 OR SICNEW = 35 OR SICNEW = 37 OR SICNEW = 38 OR SICNEW = 39) * BF17;
BF47 = (AGE > .) * BF39;
BF49 = max(0, AGE - 54.000) * BF47;
BF51 = (SICNEW = 4 OR SICNEW = 6 OR SICNEW = 7 OR SICNEW = 14 OR SICNEW = 19 OR SICNEW = 21 OR SICNEW = 23 OR SICNEW = 37) * BF21;
BF53 = (ROA > .) * BF40;
BF55 = max(0, ROA - 1.206) * BF53;
BF57 = (SICNEW = 1 OR SICNEW = 2 OR SICNEW = 3 OR SICNEW = 5 OR SICNEW = 7 OR SICNEW = 8 OR SICNEW = 15 OR SICNEW = 17 OR SICNEW = 19 OR SICNEW = 22 OR SICNEW = 23 OR SICNEW = 24 OR SICNEW = 26 OR SICNEW = 28 OR SICNEW = 29 OR SICNEW = 31 OR SICNEW = 37 OR SICNEW = 38 OR SICNEW = 39 OR SICNEW = 41) * BF1;
BF59 = (SICNEW = 12 OR SICNEW = 19 OR SICNEW = 24 OR SICNEW = 26 OR SICNEW = 30 OR SICNEW = 34) * BF16; BF61 = (SICNEW = 3 OR SICNEW = 7 OR SICNEW = 12 OR SICNEW = 19 OR SICNEW = 20 OR SICNEW = 22 OR SICNEW = 23 OR SICNEW = 25 OR SICNEW = 28 OR SICNEW = 32 OR SICNEW = 35) * BF10;
BF63 = (YRSCEO = .) * BF9;
BF64 = (YRSCEO > .) * BF9;
BF66 = max(0, 4.485668 - SALES) * BF21;
BF67 = (SICNEW = 2 OR SICNEW = 5 OR SICNEW = 6 OR SICNEW = 7 OR SICNEW = 14 OR SICNEW = 16 OR SICNEW = 21 OR SICNEW = 31 OR SICNEW = 34 OR SICNEW = 40) * BF63;
BF73 = (SICNEW = 1 OR SICNEW = 3 OR SICNEW = 4 OR SICNEW = 5 OR SICNEW = 9 OR SICNEW = 11 OR SICNEW = 12 OR SICNEW = 13 OR SICNEW = 17 OR SICNEW = 18 OR SICNEW = 19 OR SICNEW = 20 OR SICNEW = 22 OR SICNEW = 24 OR SICNEW = 27 OR SICNEW = 31 OR SICNEW = 32 OR SICNEW = 35 OR SICNEW = 37) * BF64;
BF75 = (SICNEW = 2 OR SICNEW = 4 OR SICNEW = 5 OR SICNEW = 7 OR SICNEW = 16 OR SICNEW = 18 OR SICNEW = 22 OR SICNEW = 23 OR SICNEW = 29 OR SICNEW = 30 OR SICNEW = 31 OR SICNEW = 32 OR SICNEW = 34 OR SICNEW = 38 OR SICNEW = 40 OR SICNEW = 41) * BF3;
BF77 = (YRSCEO > .) * BF6;
BF80 = max(0, 1.626 - YRSCEO) * BF77;

Estimated Equation

Y = 6.661 + 2.206 * BF1 + 0.0177346 * BF3 - 0.518625 * BF4 - 1.014 * BF5 - 0.399 * BF7 + 0.206 * BF9 - 0.203566 * BF11 - 0.439 * BF13 + 0.035 * BF17 - 0.051 * BF19 - 0.595 * BF21 + 0.019 * BF23 + 0.030 * BF24 + 0.408 * BF25 – 3.000 * BF30 - 0.00353013 * BF32 - 0.228 * BF33 + 0.035 * BF37 - 0.014 * BF38 + 0.018 * BF41 - 0.301961 * BF43 - 0.018 * BF45 - 0.052 * BF49 + 0.201 * BF51 - 0.077 * BF55 - 0.158 * BF57 - 0.869 * BF59 + 0.225 * BF61 + 0.0589534 * BF66 - 0.762 * BF67 - 0.155 * BF73 + 0.0590723 * BF75 + 0.350 * BF80;

Appendix B: Variables

AGE = age of the CEO.
NEWSIC = 4-digit SIC industry. NEWSIC is a categorical variable ranging from 1 (Broadwoven Fabric Mills, Cotton) to 41 (Services-Commercial Physical & Biological Research). See Table 3 for a complete listing of the 4-digit SIC industries.
ROA = return on assets.
SALES = sales revenue in billions of 2003 $.
Y = log of total compensation.
YEAR = observation year.
YRSCEO = years serving as CEO.