CEO Bonuses as Studied by Robust Logistic Regression

Richard P. Hauser and David Booth
Kent State University

Abstract: This investigation utilized a robust logistic regression method (BYLOGREG) to investigate CEO bonuses prior to the 2007-2009 financial crisis. The robust logistic regression analysis determined that the year and CEO tenure affected the probability that a CEO received a bonus in the 2004-2006 study period. The analysis refuted that "management entrenchment" widely influenced CEO bonus compensation because the probability of receiving a bonus was negatively related to CEO tenure. The probability of receiving of bonus declined during the 2004-2006 study period because the percentage of CEOs that received a bonus was lowest in 2006. The robust logistic regression analysis found that the current year stock return was positively and statistically significantly related to the probability that a CEO received a bonus. The analysis also showed that managerial (financial) performance in the areas of growth of sales, ROE, and growth in earnings per share increased the probability that a CEO received a bonus. In this investigation, the size of the firm and the growth rate of equity were not statistically significant. Overall, robust logistic regression correctly classified 77% of the observations on the basis of the model variables, which indicated that most CEO bonuses could be explained by firm, CEO, and financial variables.

The BY robust logistic regression proved to be robust to outliers in the CEO bonus sample studied. Interestingly, the relationship between stock return and the probability of a bonus was completely missed by a maximum likelihood (ML) logistic regression with the full CEO bonus sample, which contained outliers. After trimming the CEO bonus data set to remove outliers, the ML logistic regression coefficients changed dramatically. However, the BY robust logistic regression coefficients changed very little. Use of the residuals from the BY robust logistic equation should facilitate further inquiry into CEOs that received a bonus but were predicted to have a low probability of a bonus.

Key words: CEO compensation, data trimming, robust logistic regression.

1. Introduction

In the analysis of the 2007-2009 financial crisis, some have criticized the compensation incentives of corporate CEOs. Some shareholders believed that CEO compensation incentives led corporations to take undue risks that resulted in large equity losses. Still others believed that CEOs received bonus compensation when they should not have. While these issues and questions are highly relevant to shareholders and corporate governance, the answers to these questions are highly subjective. However, these questions did lead to the motivation for research questions that could be answered with a statistical methodology. In this study, two important research questions were investigated. The two major questions investigated were:

Research Question 1.
  What factors determine whether or not a CEO received a bonus?

Research Question 2.
  Based on available data, can we correctly predict whether or not a CEO received a bonus?

In Research Question 1, we want to learn which factors determined whether a CEO received a bonus. This question is relevant to shareholders and corporate compensation committees. Those interested parties could ensure that the CEOs interests are aligned with shareholders and that there is some measure of "pay for performance". The second research question invites solace or further inquiry. If a suitable model correctly predicted that a CEO received a bonus, shareholders and directors would feel that the CEO bonus payment was perhaps "fair" or at least explained. However, if the CEO received a bonus while the model predicted that the CEO would not receive a bonus, shareholders (and researchers) may wish to inquire further.

Based on the research questions, logistic regression seemed to be the best method of analysis. However, as shown by Pregibon (1981), logistic regression is not robust to outliers in the data. In an effort to remove the influence of outliers, financial researchers often trim the data. In this study, we investigated both trimming the data as well as robust logistic regression using the Bianco and Yohai (1996) estimator as implemented by Croux and Haesbroeck (2003).

2. Literature Review

Bianco-Yohai Robust Logistic Regression

Pregibon (1981) presented a logistic regression analysis of skin vaso-constriction data that contained outliers. His analysis showed that the maximum likelihood

(ML) estimates from logistic regression are not robust (i.e. resistant) to outliers. Kunsch et al. (1989) then proposed an Optimal Bias-Robust Estimator (OBRE) for generalized linear models. Kunsch et al. (1989) presented results for the skin vaso-constriction data, and noted that functions that excessively downweighted outlying observations lead to computational difficulties and that the estimated standard errors increased significantly. Bianco and Yohai (1996) proposed an alternative estimator that was highly robust in the logistic regression model. The Bianco and Yohai (1996) (now referred to as BY) estimator included a bounded function and a bias correction term. Croux and Haesbroeck (2003) proposed a computational method to successfully implement the BY estimator. The Croux and Haesbroeck (2003) procedure uses a bounded function to guarantee the existence of the BY estimator when the ML estimator exists and provides an algorithm to compute the BY estimate. The algorithm is available as a function in R. We utilize this BYLOGREG function to compute the BY logistic regressions in this investigation.

The Bianco-Yohai (BY) Estimator

We consider (following Bianco and Martinez (2009)) a binomial regression model where the response variable Y has a Bernoulli distribution

$$P(Y = 1|\mathbf{X} = \mathbf{x}) = F(\mathbf{x}'\boldsymbol{\beta}),$$

where F is a strictly increasing cumulative distribution function, $\mathbf{X} \in \mathbb{R}^p$ is a vector of predictor variables and $\boldsymbol{\beta} \in \mathbb{R}^p$ is the vector of unknown regression coefficients. For

$$F(t) = \frac{\exp(t)}{1 + \exp(t)},$$

the logistic regression model obtains. The maximum likelihood estimator (MLE) of $\boldsymbol{\beta}$ can be severely affected by outliers. It is known (Bianco and Martinez (2009)) that the MLE breaks down to zero for data sets containing severe outliers. This breakdown behavior has led to a number of proposals for robust estimators of $\boldsymbol{\beta}$. The robust estimator we consider is the one due to Bianco and Yohai (1996) as implemented by Croux and Haesbroeck (2003). Full details of the Bianco and Yohai (BY) estimator including robustness properties can be found in the referenced papers. Once we have the BY $\boldsymbol{\beta}$, we wish to test hypotheses about the components of $\boldsymbol{\beta}$. Bianco and Martinez (2009) show that using BY estimates in a Wald-type statistic yields an asymptotic central Chi-square distribution as does the classical Wald statistic. Full details of the above are given in Sections 2, 3, 4, and 5 of Bianco and Martinez (2009). The reason BY estimates were used in this study was first, all the estimates are robust (Bianco and Yohai (1996))

and with the recent work of Bianco and Martinez (2009) we know the asymptotic distribution of the corresponding Wald-type statistics.

Wald-Type Inference Tests

The Wald inference test is applied to tests of hypotheses on individual logistic regression coefficients in the logistic regression model. The objective was to test the null hypothesis

$$H_0 : \beta_j = 0,$$

against the alternative hypothesis that

$$H_1 : \beta_j \neq 0.$$

Myers et al. (2002) showed that for ML logistic regression coefficient estimators, the Wald test statistic was

$$WaldChi - square = \left[\frac{estimated\ coefficient}{standard\ error\ of\ coefficient} \right]^2,$$

and is asymptotically distributed as a Chi-square under the null hypothesis. This test is shown on standard computer output with software packages such as SAS along with the corresponding p values for each coefficient hypothesized in the logistic regression model. Given the properties of the BY logistic estimators, we computed the Wald Chi-square test statistic, now strictly speaking a Wald-type test statistic, for the estimates from the BY logistic regression with the same equation. Then, p-values for the BY logistic regression coefficients were computed from Chi-square distribution tables.

Further, such Wald-type tests work for the following reasons. Because the BY estimator is robust (Bianco and Martinez (2009); Croux and Haesbroeck (2003); Bianco and Yohai (1996)), it is resistant to deviations from the standard logistic regression assumptions and hence should satisfy the usual assumptions for logistic regression better than an MLE alone (which is not resistant) for logistic regression when outliers are present in the data. Relevant theorems can be found in the paper by Bianco and Martinez (2009), which discusses in detail Wald-type tests for the BY estimator. In particular, they show that these statistics are asymptotically Chi-square as was the original Wald statistic and hence can be used as described above.

Goodness of fit tests

The logistic regression goodness of fit tests required a new measure when robust logistic regression was considered. Kordzakhia et al. (2001) determined that

the Chi-square arcsine transformation satisfied model fit criteria and was used by Croux and Haesbroeck (2003), where the Chi-square arcsine transformation was defined as

$$\chi^2_{\mathrm{arc}} = \Sigma 4[\arcsin \sqrt{y} - \arcsin \sqrt{\pi}]^2,$$

where π are the fitted probabilities.

3. Data and Research Methodology

CEO Bonus Data

Sample

Our initial sample of CEO bonus pay data consisted of CEOs from the Standard & Poor's ExecuComp database. The ExecuComp database provides annual data on executive compensation for most firms in the Standard & Poor's 1500 Index. The CEO bonus pay, variable (bonus), includes only cash bonus compensation. We utilized Standard & Poor's Compustat database for financial data on the firms. In order to evaluate corporate behavior prior to the financial crisis in 2007-2008, we included the years 2004-2006 in the study.

After merging the data extracted from the ExecuComp and Compustat databases, we removed any observation with missing data. We chose not to impute missing data because that adds an additional variable to the estimator comparison. If the method comparison was not the main goal of this study, imputation may have been used. The concern with the use of imputed values and the required errors in variables regression was that these procedures would complicate the desired study objective of method comparison. Furthermore, imputation is not commonly used in similar studies in the finance literature.

To eliminate any potential effects of signing bonus data, we restricted the sample to CEOs with a tenure greater than one year. Since we consider only experienced CEOs, we further restricted the sample in year 2005 to CEOs with two years of tenure to ensure a consistent CEO base. Likewise in year 2006, we restricted the sample to CEOs with a three years of tenure in order to eliminate companies with CEO turnover. The size of the obtained data set was 3,685 observations.

Variables

A primary purpose of the study was to determine if CEO bonus pay, in the year considered, depended on the managerial performance variables of return on shareholders' equity, growth in earnings per share, growth in equity, and growth

in sales. In order to determine this, the following variables were extracted from the Compustat database: (ROE) or return on shareholders' equity, (geps) or growth in earnings per share, (geq) or growth in equity, and (gsales) or growth in sales. The growth rate data were based on the one-year rates. For the control variable on company size, we followed the procedure of Cordeiro and Veliyath (2003) and Core et al. (2008) and used the variable (sales), or net annual sales. Other control variables included (tenure) for the CEO tenure and (ret1yr) for the current year stock total return. Based on prior research, logarithms of sales and tenure were used in the regressions. Plots of the 2004-2006 data indicated yearly fluctuations, and the variable (year) was included. The bonus response variable then was included in the analysis and was set to a value of 1 if the CEO obtained a bonus, otherwise to zero. This bonus response variable becomes important in the study because we used logistic regression to determine if managerial performance variables played a part in the decision to pay a CEO a bonus.

Univariate Statistics

Inspection of the summary statistics revealed that the data set had several problems common to financial data sets. First, the presence of outliers was noted by inspection of the maximum values compared to the mean values. Another problem detected in the data set was the large spread between the mean and median values for several variables also indicating likely outliers. From our preliminary data plots, the percentage of CEOs that received a bonus varied on a year to year basis. The summary statistics also indicated that CEOs who received a bonus had higher mean and median stock returns, return on equity, growth in equity, growth in earnings per share, and growth in sales.

4. Empirical Logistic Regression Results

Logistic Regression

While the univariate results indicated that there seems to be relationship between a firm's financial results and the decision to award the CEO a bonus, the relationship needs to be confirmed with regression. To analyze the decision to award a CEO a bonus, we utilized a logistic regression. As Cody and Smith (2006) describe, a logistic regression uses a logit transformation that forces the prediction equation to yield values between zero and one. This situation is very well suited to model the bonus response variable that can only take on values of 1 if the CEO obtained a bonus or zero otherwise. The logistic regression equation is then interpreted as predicting the natural log of the odds of a CEO receiving a bonus or not. In order to analyze the effects of the firm's performance

variables, we developed a logistic regression equation with the control variables and financial performance variables as:

$$\ln(\text{odds of CEO receiving a bonus}) = \beta_0 + \beta_1 \ln(\text{sales}) + \beta_2 \ln(\text{tenure}) + \beta_3 \text{year}$$
$$+ \beta_4 \text{ret1yr} + \beta_5 \text{ROE} + \beta_6 \text{geps} + \beta_7 \text{geq} + \beta_8 \text{gsales} + \varepsilon. \qquad \text{(Model 1)}$$

We report the results of the ML logistic regression in Table 1, and again note that these coefficients are maximum likelihood estimates. The Pregibon (1981) regression diagnostic plots are presented in Figure 1. From Table 1, Research Question 1 can be answered. The control variables ln(sales) and year are significant with p values of .0120 and $< .0001$ respectively, but ln(tenure) and stock return are not significant. Based on these results, one would have concluded that the probability that a CEO received a bonus depended on the size of the company and the year but not on the CEO's tenure or the firm's current year stock return. In terms of the financial performance variables, return on equity, growth in earnings per share, and growth in sales are all significant at the 5% level. The growth in firm equity was not significant. It appears that the first research question is answered. That is, after controlling for some firm level and CEO variables, financial performance variables return on equity, growth in earnings per share, and growth in sales all statistically increased a CEO's probability of receiving a bonus. The logistic regression prediction equation correctly classified 79% of the observations. While these results seem to address the research question, analysis of the regression diagnostic plots indicated problems. The regression diagnostic plots shown in Figure 1 indicated outliers and one specific outlier observation that was very influential on the regression as indicated by the extremely large Cook's distance statistic. Thus, these regression results are questionable based on the results of Pregibon (1981) and further analysis is required before we can safely answer Research Question 1.

Robust Logistic Regression

Because of the outliers in the CEO bonus data set and their impact on the ML logistic regression, Model 1 was analyzed with robust (i.e. outlier resistant) logistic regression. Using the Bianco and Yohai (1996) estimator as implemented by Croux and Haesbroeck (2003), we obtained the BY robust logistic regression results presented in Table 2. For some variables, the estimated parameters and p values from the BY robust logistical regression are very different from the ML regression. For the BY robust logistic regression, the control variables ln(tenure), year, and 1-yr stock return were significant, but the ln(sales) was not significant. The financial performance variables return on equity and growth of sales were significant, but the growth in earnings per share and growth in equity were not

Table 1: ML Logistic Regression of CEO bonus data

| Parameter	Estimate	Standard Error	Wald Chi-square	Pr>Chisq
Intercept	2569.4	110.4	542.07	< .0001
ln(sales)	0.0659	0.0262	6.3178	0.012
ln(tenure)	-0.0693	0.0583	1.4132	0.2345
Year	-1.2813	0.0551	541.61	< .0001
ret1yr	0.000071	0.0002	0.1125	0.7373
ROE	0.00344	0.0011	9.1272	0.0025
Geps	0.00024	0.00009	7.3648	0.0067
Geq	-0.00043	0.0006	0.6064	0.4361
Gsales	0.0098	0.002	23.194	< .0001
Classification % Correct	78.6%			
Goodness of fit Chi-Square Arcsine	6206.5			
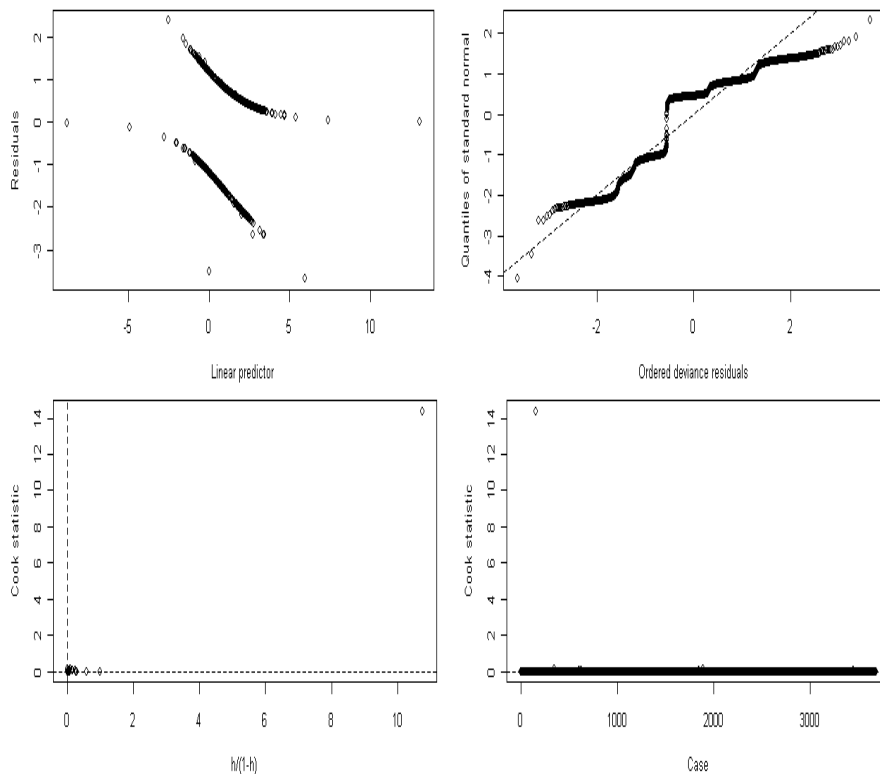
Figure 1: ML Logistic Regression Diagnostic Plots for CEO Bonus Data

significant at the 5% level. Overall, the BY robust logistic model had a slightly lower correct classification at 77% and a slightly worse chi-square arcsine goodness of fit.

Table 2: BY Robust Logistic Regression of CEO bonus data

Parameter	Estimate	Standard Error	Wald Chi-square	Pr>Chisq
Intercept	2318.7	62.626	1370.82	$< .0001$
ln(sales)	0.0000934	0.016	0.000034	0.9953
ln(tenure)	-0.07772	0.0331	5.513	0.0189
Year	-1.1561	0.0312	1369.87	$< .0001$
ret1yr	0.006414	0.0011	32.796	$< .0001$
ROE	0.011894	0.0050	5.704	0.0169
Geps	0.002246	0.00142	2.519	0.1124
Geq	0.006642	0.0080	0.681	0.4092
Gsales	0.007939	0.0018	19.410	$< .0001$
Classification %Correct	77%			
Goodness of fit Chi-Square Arcsine	6348.3			

While the overall fit of the model was much the same using the BY robust (outlier resistant) logistic regression algorithm and ML logistic regression, the interpretation of the effect of many variables in determining the probability of CEO receiving a bonus was very different. The comparison between the BY robust and ML logistic regression is summarized in Table 3. In the ML logistic regression, the coefficients for ln(sales), year, ROE, geps, and gsales are significant at the 5% level. In the BY robust logistic regression, the coefficients for ln(tenure), year, ret1yr, ROE, and gsales are significant at the 5% level. The difference in interpretation of the effect results from either the size of the estimated coefficient or the standard error. For example with the ML logistic regression, the value of the estimated coefficient on the stock return variable ret1yr is 0.00007. This value is almost a factor of 100 times smaller than the estimated coefficient on the variable ret1yr estimated by BY logistic regression, which leads to a different interpretation. Meanwhile other estimated coefficients had similar values, but very different standard errors. For example, the estimated coefficient on the ln(tenure) variable is -0.069 for ML logistic regression and -0.078 for BY logistic regression. However, the standard error for ML logistic regression is almost twice the standard error from BY logistic regression, making the variable ln(tenure)

only significant with the BY logistic regression. The differences in interpretation are caused by the presence of outliers in the sample data. We know from the summary descriptive statistics that the data set contains outliers, and we also know that the presence of outliers has a tremendous influence on the ML logistic regression as seen by the regression diagnostic plots in Figure 1. Finally, the outliers in the data set cause the differences in interpretation of the regressions.

Table 3: Comparison of Logistic Regressions of CEO bonus data

Parameter	ML Estimate		BY Robust Estimate	
Intercept	2569.4	*	2318.7	*
	($< .0001$)		($< .0001$)	
ln(sales)	0.0659	*	0.0000934	
	(0.012)		(0.9953)	
ln(tenure)	-0.0693		-0.07772	*
	(0.2345)		(0.0189)	
Year	-1.2813	*	-1.1561	*
	($< .0001$)		($< .0001$)	
ret1yr	0.000071		0.006414	*
	(0.7373)		($< .0001$)	
ROE	0.00344	*	0.011894	*
	(0.0025)		(0.0169)	
Geps	0.00024	*	0.002246	
	(0.0067)		(0.1124)	
Geq	-0.00043		0.006642	
	(0.4361)		(0.4092)	
Gsales	0.0098	*	0.007939	*
	($< .0001$)		($< .0001$)	

p value in parentheses
* significant at 5% level

Based on the results in Table 3, we summarize below how the following variables affected the probability that a CEO received a bonus in the 2004-2006 period before the financial crisis.

Company size:

The BY robust logistic regression indicated that the effect of company size as proxied by ln(sales) was statistically insignificant to the probability that a CEO received a bonus.

CEO tenure:

The BY logistic robust regression indicated that the effect of the CEO's tenure as proxied by ln(tenure) was statistically significant but negatively related to the

probability that a CEO received a bonus. This indicated that the management entrenchment did not widely influence CEO bonus over the 2004-2006 period prior to the financial crisis.

Year:

The BY robust logistic regression indicated that the effect of the year was statistically significant. The negative coefficient indicated that the probability of CEO bonus was lower at the end of the study period, 2006. Note that Table 2 showed a dramatic decline in the percentage of CEO's that received a bonus in 2006.

Stock Return:

The BY robust logistic regression indicated that the effect of the current year stock return was statistically significant and that a high stock return improved the CEO's probability of receiving a bonus in the study period. It should be noted that the ML logistic regression missed this relationship.

Managerial Performance:

The BY robust logistic regression indicated that growth in sales and ROE positively and statistically significantly improved a CEO's probability of receiving a bonus in the study period. The growth in sales was most significant, while the growth in equity was not significant.

Trimmed CEO Bonus Sample

An approach that is commonly used in the financial literature has been to trim the sample to remove outliers. In order to see the effect of such trimming, we trimmed the CEO bonus data set in the following manner:

Variable	Range to remove
stock ret1yr	< -100% and >100%
ROE	< -100% and >100%
geq	< -100% and >100%
geps	< -100% and >100%
gsales	< -100% and >100%

The summary statistics for the trimmed CEO bonus data set are given in Table 4 below. While trimming partially achieves the desired effect of removing some extreme data points and moving the mean and median closer together, it comes at the expense of trimming about 900 observations from the data set. Both ML logistic regression and BY robust logistic regression were performed on the trimmed CEO bonus data set and the results are summarized in Table 5. With

the trimmed CEO bonus data set, the results of the ML logistic regression and BY robust regression are quite similar. While the BY robust logistic regression results changed little from the full sample and provided a slightly better goodness of fit, the ML logistic regression results changed substantially. Furthermore, the standard trimming approach does not take into account multivariate outliers while the BY method does.

Table 4: Trimmed 2004 - 2006 Sample

Variable	Units	N	Mean	Median	Std Dev	Min	Max
CEO cash bonus	Thousands $	2797	980.67	440.6	1949.4	0	310000
CEO tenure	Years	2797	9.3	6.85	7.56	1	55.8
Firm sales	Millions $	2797	6844.3	1594.9	19553	1.89	328213
Firm 1-yr stock market total return	%	2797	12.13	11.21	25.87	-73.88	99.21
Firm return on equity (ROE)	%	2797	13.04	12.75	13.97	-99.3	95.91
Firm 1-yr equity growth rate	%	2797	10.63	10.72	13.33	-83.83	93.04
Firm 1-yr earnings per share growth rate	%	2797	13.08	13.77	37.04	-100	100
Firm 1-yr sales growth rate	%	2797	13.74	11.28	16.26	-90.34	99.67
Bonus Response	-	2797	0.71	1	0.45	0	1
ln (sales)	ln(Millions $)	2797	7.51	7.37	1.55	0.64	12.7
ln(tenure)	ln(years)	2797	1.96	1.92	0.74	0.006	4.02
Year	-	2797	2004.9	2005	0.81	2004	2006

In the ML logistic regression, the coefficients for year, ret1yr, and gsales are now significant at the 5% level. In the BY robust logistic regression, the coefficients for ln(tenure), year, ret1yr, ROE, geps, and gsales are significant at the 5% level. Removing some of the outliers makes the results from the ML logistic regression approach the BY logistic regression. For example, the estimated coefficient on the stock return variable ret1yr from ML regression is now 0.009 with the trimmed data set. The estimated coefficient on the stock return variable ret1yr from BY regression is a very similar 0.008; whereas before (with the full sample) the value from the ML regression was almost 100 times smaller. Since removing some of the outliers makes the results of ML logistic regression approach the BY logistic regression, we rely on the results from the outlier resistant, robust BY logistic regression.

Research Question 2

Table 5: Comparison of Logistic Regressions of Trimmed CEO bonus data

Parameter	ML Estimate		BY Robust Estimate	
Intercept	2741.2	*	2828.1	*
	(< .0001)		(< .0001)	
ln(sales)	0.0161		-0.02438	
	(0.6150)		(0.1604)	
ln(tenure)	-0.1247		-0.1157	*
	(0.0709)		(0.0026)	
Year	-1.3668	*	-1.410	*
	(< .0001)		(< .0001)	
ret1yr	0.00924	*	0.00802	*
	(< .0001)		(< .0001)	
ROE	0.00548		0.0179	*
	(0.460)		(0.0001)	
Geps	0.00231		0.00437	
	(0.0881)		(< .0001)	*
Geq	0.0131		0.0022	
	(0.103)		(0.6468)	
Gsales	0.0085	*	0.0114	*
	(0.0067)		(< .0001)	
Classification %Correct	77.70%		77.70%	
Goodness of fit Chi-Square Arcsine	4575.7		4545.4	

p value in parentheses
* significant at 5% level

To answer Research Question 2, we need the prediction equation that answered Research Question 1. The prediction equation used to answer Research Question 1 indicated that the logistic regression model classified about 80% of the observations correctly. The investigation of the BY robust logistic regression indicated that it was a procedure that was truly robust to the outliers, thus the further investigation for Research Question 2 utilized the BY robust logistic regression prediction equation. While strictly speaking prediction accuracy should be studied by cross validation, we are here only interested in the residuals from the fitting step as discussed below.

Given that the BY robust logistic regression prediction equation classified about 80% of the observations correctly, it was actually more interesting to look

at cases where the prediction equation did not predict correctly. We are most interested in the situation where the CEO received a bonus, but the prediction equation predicted a low probability of a CEO bonus. To analyze these cases, we computed the residuals, defined as the actual bonus response for the observation minus the predicted probability from the prediction equation. These residuals were plotted and displayed in Figure 2. The residuals then of most interest were the observations where the residual was nearly 1, which corresponds to cases where the CEO received a bonus but the prediction equation predicted zero (or very low) probability of a bonus. We sorted the data set by residual and extracted the 10 observations with the largest residuals. These observations are summarized in Table 6.
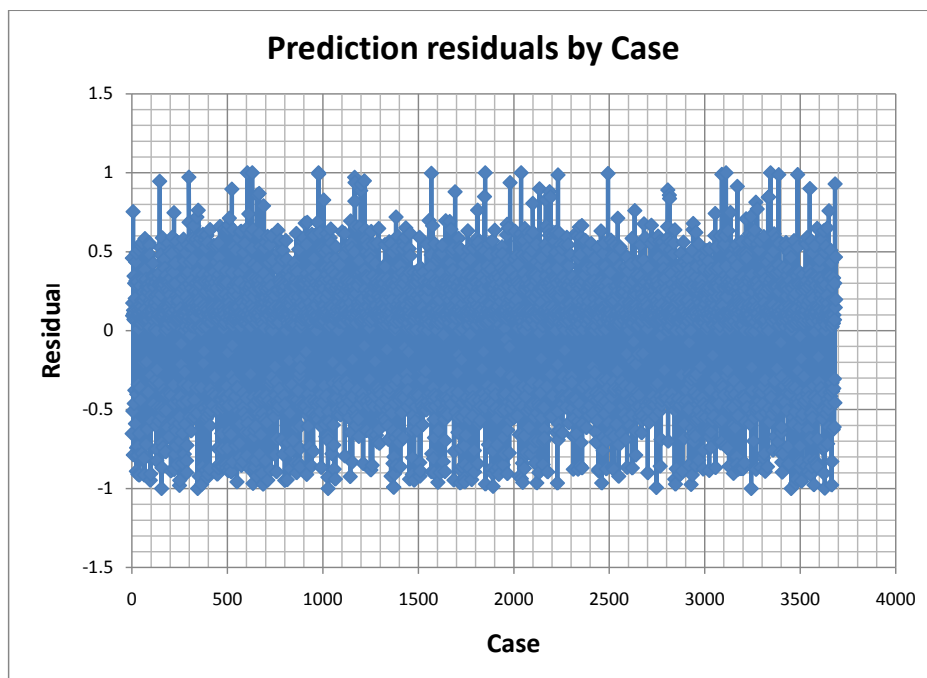

Figure 2: Prediction residuals for the CEO bonus data set

Table 6 reveals why the prediction equation failed to predict that these observations would be cases where the CEO received a bonus. Although not all the financial variables are listed, inspection of the stock return and return on equity seemed to suffice. Shareholders of firms that had such poor financial performance would have reason to question the CEO bonus in these cases. Investigating all of these cases was beyond the scope of this investigation, but such a methodology would be of interest to shareholders and corporate directors. More importantly, it should be noted that most of the 10 extreme observations would have been lost

Table 6: The 10 observations with the largest residuals

Case	Bonus	BYpred	Residual	Ret1yr %	ROE %	Bonus (thousand$)
604	1	0.0000	1.0000	7.8	-6.6	182.8
630	1	0.0000	1.0000	-16.3	-9.5	238
2038	1	0.0000	1.0000	-41.5	-129.1	135
977	1	0.0001	0.9999	-24.7	-1027.7	150
3111	1	0.0002	0.9998	33.3	-4.3	204.1
3344	1	0.0002	0.9998	23.9	-14.9	8046
1850	1	0.0003	0.9997	-43.9	-688.0	35
1568	1	0.0033	0.9967	11.1	-255.4	108.4
2493	1	0.0046	0.9954	-31.3	-426.6	143.2
978	1	0.0083	0.9917	-26.4	-209.4	250

due to trimming. Therefore, trimming the data set would have removed cases that would be of interest for investigation. Using the BY logistic regression on the full data set, we are able to identify all residuals that may be of interest for further investigation.

5. Limitations and Future Research

This study used the Bianco and Yohai (1996) estimator as implemented by Croux and Haesbroeck (2003). Croux and Haesbroeck (2003) mention that S PLUS includes other robust logistic regression estimators such as the CUBIF and MALLOWS estimators. Further research might compare these estimators with the BY robust logistic regression estimator used.

An additional area of research is to look at small sample properties of both the BY robust estimators and ML estimators in logistic regression. There are many open questions in this area.

Further research on the residuals where the CEO received a bonus but the prediction equation predicted a low probability may shed insights for better "pay for performance" incentive compensation.

6. Conclusions

This study used a new methodology for investigating CEO bonus compensation using a data set that contained outliers. The results are important to shareholders and corporate directors that set CEO incentive compensation. The BY robust logistic regression model determined that the following variables affected the probability that a CEO received a bonus in the 2004-2006 period before the financial crisis.

Company size:

The BY robust logistic regression indicated that the effect of company size as proxied by ln(sales) was statistically insignificant in both the full CEO bonus sample and the trimmed sample.

CEO tenure:

The BY logistic robust regression indicated that the effect of the CEO's tenure as proxied by ln(tenure) was statistically significant but negatively related to the probability that a CEO received a bonus. This indicated that the management entrenchment did not widely influence CEO bonus over the 2004-2006 period prior to the financial crisis.

Year:

The BY robust logistic regression indicated that the effect of the year was statistically significant. The negative coefficient indicated that the probability of CEO bonus was lower at the end of the study period, 2006. Table 2 showed a dramatic decline in the percentage of CEO's that received a bonus in 2006. Since 2006 CEO bonuses were most likely paid in the first quarter of 2007, it was likely that corporations were already anticipating some financial difficulty in 2007.

Stock Return:

The BY robust logistic regression indicated that the effect of the current year stock return was statistically significant and that a high stock return improved the CEO's probability of receiving a bonus in the study period. It should be noted that the ML logistic regression missed this relationship with the full CEO bonus sample, which contained large outliers.

Managerial Performance:

The BY robust logistic regression indicated that growth in sales, growth in earnings, and ROE positively and statistically significantly improved a CEO's probability of receiving a bonus in the study period. The growth in sales was most significant, while the growth in equity was not significant.

The BY logistic regression prediction equation correctly classified 77% of the observations, which indicated that probability that a CEO received a bonus could be fairly well explained by firm and CEO control variables as well as the firm's financial performance. This indicated that most CEO bonuses were "fair" or at least explainable. Of interest, were those CEOs that received a bonus, but the prediction equation indicated a very low probability of bonus. A review of these cases indicated that further inquiry was indeed justified. In the 10 largest residual observations, CEOs received bonus compensation even when the ROE

was negative.

In a comparison of the logistic regression methods, the BY robust logistic regression indicated that it was indeed robust to outliers in the data set. The BY robust logistic regression results changed rather little between the full CEO bonus sample and the trimmed CEO bonus sample. The only significant change was that the growth in earnings became significant in the BY logistic regression on the trimmed CEO bonus sample. Meanwhile, the ML logistic regression clearly showed problems dealing with outliers in the full CEO bonus sample. The regression diagnostic plots indicated the outliers and the coefficients were largely different from the BY logistic regression coefficients. Using the trimmed CEO data set which removed many outliers, the ML logistic regression coefficients more closely matched the BY logistic regression coefficients. It was clear that outliers in the data set affected the ML logistic regression; furthermore, these outliers caused a different interpretation of the ML logistic regression results. The BY logistic regression was robust to outliers and enabled analysis of residual observations that would have been missed by trimming.

Acknowledgement

References

Bianco, A. M. and Martinez, E. (2009). Robust testing in the logistic regression model. *Computational Statistics and Data Analysis* **53**, 4095-4105.

Bianco, A. M. and Yohai, V. J. (1996). *Robust Estimation in the Logistic Model, in Robust Statistics, Data Analysis, and Computer Intensive Methods*, 17-34; Lecture Notes in Statistics 109, Springer Verlag, New York.

Cody, Ronald P. and Smith, J. K. (2006). *Applied Statistics and the SAS programming language.* 5th ed. Pearson Prentice Hall, New Jersey.

Corderio, J. and Veliyath, R. (2003). Beyond pay for performance: A panel study of the determinants of CEO compensation. *American Business Review* **21**, 56-66.

Core, J., Holthausen R. and Larker, D. (1999). Corporate governance, chief executive officer compensation and firm performance. *The Journal of Financial Economics* **51**, 371-406.

Core, J., Guay, W. and Larker, D. (2008). The power of the pen and executive compensation. *The Journal of Financial Economics* **88**, 1-25.

Croux, C. and Haesbroeck, G. (2003). Implementing the Bianco and Yohai estimator for Logistic Regression. *Computing and Statistical Data Analysis* **44**, 273-295.

Kordzakhia, N., Mishra, G. D. and Reiersolmoen, L. (2001). Robust estimation in the logistic regression model. *Journal of Statistical Planning and Inference* **98**, 211-223.

Kunsch, H.R., Stefanski, L. A. and Carroll, R. J. (1989). Conditionally unbiased bounded Influence estimation in general regression models with applications to generalized linear models. *Journal of the American Statistical Association* **84**, 460-466.

Myers, R. H., Montgomery, D. C. and Vining, G. G. (2002). *Generalized Linear Models*. John Wiley & Sons, New York.

Pregibon, D. (1981). Logistic Regression Diagnostics. *The Annals of Statistics* **9**, 705-724.

Richard P. Hauser
Department of Finance
Kent State University
P.O. Box 5190 Kent, OH 44242-0001
rhauser1@kent.edu

David Booth
Department of Management and Information Systems
Kent State University
P.O. Box 5190 Kent, OH 44242-0001
dbooth@kent.edu