Predicting Bankruptcy with Robust Logistic Regression

-Richard P. Hauser and David Booth

Introduction:

The prediction of corporate bankruptcy is an important and widely studied topic. Creditors and investors in corporations need to be able to predict the probability of default for profitable business decisions. For banks, accurate assessment of the probability of bankruptcy can lead to sounder lending practices as well as better fair value estimates of interest rates that reflect credit risks. 0 and 1. However, two major problems with logistic regression occur due to the nature of the bankruptcy problem. First, we assert that bankrupt corporations can be viewed as outliers from the perspective of a group of healthy firms . Thus in the case of corporate bankruptcy, it is possible to get rather high percentages correctly predicted even when the models' prediction of bankruptcy is extremely poor.

Literature review:

The maximum likelihood (ML) estimates from logistic regression are not resistant to outliers.

Bianco-Yohai Robust Logistic Regression:

Bianco and Yohai (1996)proposed an alternative estimator that was highly robust in the logistic regression model.We utilize BYLOGREG function to compute the BY logistic regressions in this investigation.

The Bianco-Yohai (BY) Estimator:

Consider a binomial regression model where the response variable Y has a Bernoulli distribution: $\mathbf{P(Y = 1 \mid X = x) = F(x'\beta)}$

where F is a strictly increasing cumulative distribution function,

$$\mathbf{F(t) = exp(t)/1 + exp(t)}$$

Bianco and Martinez (2009) show that using BY estimates in a Wald-type test statistic yields an asymptotic central Chi-square distribution as the test statistic's sampling distribution, just as does the classical Wald statistic in the ML case. Thus for the inference tests, the quadratic form of the Wald-type test statistic reduces to :

$$\mathbf{z_i^2 = [\hat{\beta}_i/\text{Standard error of } \hat{\beta}_i]^2}$$

The reason BY estimates were used in this study was rst, all the estimates are robust to outlier. Finally with the recent work of Bianco and Martinez (2009), we can make inference tests .

Data and Methodology :

Data and Variables :

Financial institutions such as commercial banks and investment banks are excluded from the data set because these financial institutions can be affected by actions of government regulators. Furthermore, financial ratio information was required for the bankrupt firms. The financial data was extracted from COMPUSTAT. COMPUSTAT financial data was extracted for the bankrupt firms for 2006 and 2007, which corresponded to the two-year and one-year periods, respectively, prior to bankruptcy filing.

Each firm was described by Altman's five financial ratios since the prediction capabilities of these ratios are well documented in the prior literature

1. WCTA = working capital / total assets.
2. RETA = retained earnings / total assets.
3. EBITTA = earnings before interest and taxes / total assets.
4. MEDEBT = market value of equity / book value of total debt.
5. SALETA = sales / total assets.

As the firms approached filing for bankruptcy, the WCTA, RETA, and EBITTA ratios all decreased, which indicated the bankrupt firms were in a worsening financial condition as they approached filing for bankruptcy. Predicting Bankruptcy with Robust Logistic Regression 571 prior to filing for bankruptcy, the firms had lower market values for equity, high total debt, or both. The firm with the lowest WCTA, RETA, and EBITTA ratios is from the non-bankrupt firm sample.

<u>Cross - validation Technique :</u>

The cross-validation technique enables us to use the whole data set so that any bias effect would be minimized . The total data set consists of 24 firms that filed for bankruptcy in 2008-2009 and 48 firms that did not file for bankruptcy for a total data set of 72 firms. As shown in Table 1, the total data set is divided into 3 equal and mutually exclusive subsets. The prediction equation developed from the training set is then used to predict the probability of bankruptcy for firms in subset 3-the testing data set.

<u>Logistic Regression Results</u>

In BY robust logistic regression three fold cross validation is used to compare the classification and the prediction of BY estimator with the ML Logistic Regression.

		SUBSET		
		1	2	3
RUN	1	Training 16 non-bankrupt 8 bankrupt firms	Training 16 non-bankrupt 8 bankrupt firms	Testing 16 non-bankrupt 8 bankrupt firms
	2	Testing 16 non-bankrupt 8 bankrupt firms	Training 16 non-bankrupt 8 bankrupt firms	Training 16 non-bankrupt 8 bankrupt firms
	3	Training 16 non-bankrupt 8 bankrupt firms	Testing 16 non-bankrupt 8 bankrupt firms	Training 16 non-bankrupt 8 bankrupt firms

Each subset consists of 16 randomly selected non-bankrupt firms and 8 randomly selected bankrupt firms for a total of 24 firms.

Table 2 and 3 summarises the classifications and predictions as applied to the three-fold cross-validation scheme with the 2006 data (Table 2) and 2007 data (Table 3) for both ML logistic regression and BY robust logistic regression.

ML Logistic Regression	Training Set			Testing Set		
	Non-bankrupt	Bankrupt	Overall	Non-bankrupt	Bankrupt	Overall
Run 1						
Correct #	30	9	39	12	2	14
Total #	32	16	48	16	8	24
Percent Correct	93.75%	56.25%	81.25%	75.00%	25.00%	58.33%
Run 2						
Correct #	32	0	32	16	0	16
Total #	32	16	48	16	8	24
Percent Correct	100.00%	0.00%	66.67%	100.00%	0.00%	66.67%
Run 3						
Correct #	30	6	36	12	1	13
Total #	32	16	48	16	8	24
Percent Correct	93.75%	37.50%	75.00%	75.00%	12.50%	54.17%
Summary						
Correct #		15	107		3	43
Total #		48	144		24	72
Percent Correct		31.25%	74.31%		12.50%	59.72%

BY Logistic Regression	Training Set			Testing Set		
	Non-bankrupt	Bankrupt	Overall	Non-bankrupt	Bankrupt	Overall
Run 1						
Correct #	28	10	38	12	4	16
Total #	32	16	48	16	8	24
Percent Correct	87.50%	62.50%	79.17%	75.00%	50.00%	66.67%
Run 2						
Correct #	28	11	39	14	3	17
Total #	32	16	48	16	8	24
Percent Correct	87.50%	68.75%	81.25%	87.50%	37.50%	70.83%
Run 3						
Correct #	30	6	36	12	1	13
Total #	32	16	48	16	8	24
Percent Correct	93.75%	37.50%	75.00%	75.00%	12.50%	54.17%
Summary						
Correct #		27	113		8	46
Total #		48	144		24	72
Percent Correct		56.25%	78.47%		33.33%	63.89%

ANALYSIS OF REGRESSION COEFFICIENT

PANEL A 2006 Financial Data

	Bankruptcy Prediction					
	ML Logistic Regression			BY Logistic Regression		
	RUN 1	RUN 2	RUN 3	RUN 1	RUN 2	RUN 3
Probability	0.448	0.329	0.268	0.994	0.33	0.278
Correct Prediction #	0	0	0	1	0	0
OVERALL	ML Logistic Regression			BY Logistic Regression		
Correct Prediction %	0.0%			33.3%		

PANEL B 2007 Financial Data

	Bankruptcy Prediction					
	ML Logistic Regression			BY Logistic Regression		
	RUN 1	RUN 2	RUN 3	RUN 1	RUN 2	RUN 3
Probability	0.452	0.396	0.32	0.661	0.708	0.326
Correct Prediction #	0	0	0	1	1	0
OVERALL	ML Logistic Regression			BY Logistic Regression		
Correct Prediction %	0.0%			66.7%		

Note that predicted probabilities greater than 0.5 were considered correct, and less than 0.5 were considered incorrect.

ANALYSIS OF DEVIANCE RESIDUALS

PANEL A 2006 Financial Data

	Parameter Estimates (Standard errors below in parentheses)					
	ML Logistic Regression			BY Logistic Regression		
Variable	RUN 1	RUN 2	RUN 3	RUN 1	RUN 2	RUN 3
Intercept	-0.2387	-0.4033	-0.6379	1.4597	-0.4014	-0.6079*
	(.7219)	(.5832)	(.6253)	(.9489)	(.2998)	(.3426)
WCTA	0.0603	-1.7855	-1.8254	22.0113**	-1.7773*	-1.7395*
	(1.5791)	(1.7376)	(1.9722)	(10.6710)	(.9333)	(.9656)
RETA	0.0762	0.2544	0.8465	-0.5861	0.2532**	0.8067***
	(.1416)	(.2240)	(.6722)	(.3172)	(.1103)	(.2971)
EBITTA	-0.0165	-0.2476	-1.6378	-1.4556	-0.2465	-1.5607*
	(1.3017)	(1.5305)	(2.1121)	(1.0454)	(.7112)	(.9113)
MEDEBT	-0.0697	-0.0006	-0.0006	-1.8140**	-0.0006	-0.0006
	(.0584)	(.0013)	(.0015)	(.74122)	(.0005)	(.0005)
SALETA	0.2587	0.1734	0.5077	0.7750	0.1727	0.4838*
	(.4913)	(.4189)	(.4818)	(.5796)	(.2300)	(.2635)

PANEL B 2007 Financial Data

	Parameter Estimates (Standard errors below in parentheses)					
	ML Logistic Regression			BY Logistic Regression		
Variable	RUN 1	RUN 2	RUN 3	RUN 1	RUN 2	RUN 3
Intercept	-0.1635	-0.4183	-0.4915	0.9542	0.4508	-0.4733
	(.7233)	(.5592)	(.6087)	(.6520)	(.3509)	(.3150)
WCTA	-0.9024	-0.0858	-2.0125	-2.9568**	3.1891	-1.9381**
	(.7520)	(.6510)	(1.7516)	(1.4804)	(2.0803)	(.8633)
RETA	0.0827	0.0580	0.8992	0.1301**	-0.4185*	0.8659***
	(.1288)	(.1697)	(.6650)	(.0636)	(.2383)	(.3303)
EBITTA	0.9285	-0.0315	-0.9785	1.8078**	1.1725	-0.9423
	(1.2619)	(1.1177)	(1.9136)	(.7709)	(1.2112)	(.9276)
MEDEBT	-0.1543	-0.0208	-0.0162	-0.6442*	-1.0838*	-0.0156
	(.0903)*	(.0212)	(.0240)	(.3357)	(.5699)	(.0128)
SALETA	0.3658	0.0956	0.4447	0.1441	0.3329	0.4282**
	(.5262)	(.3707)	(.4474)	(.3217)	(.3247)	(.2126)

* indicates asymptotic significance at 10% level
** indicates asymptotic significance at 5% level
*** indicates asymptotic significance at 1% level

Conclusion:

BY robust logistic regression is robust to the presence of outliers, we showed evidence that BY robust logistic regression improves on the ML logistic regression when outliers are present in the sample. Our analysis indicates that if the BY robust logistic regression significantly changes the estimated regression coefficients from ML logistic regression, then the BY robust logistic regression method can significantly improve the classification and prediction of bankrupt firms. This is strong evidence that BY robust logistic regression should be used as a robustness check on ML logistic regression. If a difference exists, BY robust logistic regression should be used as the primary classifier of bankrupt firms.

Submitted by:

ANANDAVARMAN K S(19S003),
BHUMIKA K(19S006),
CHANDRADHITHYA K G(19S007),
SHIBANI K R(19S033),
SHOBIKA R(19S034)