Survival Analysis Applied to Business Failure during the 2007 Recession

Abstract

This study applies survival analysis to business failure cases of publicly traded firms during the first five quarters of the 2007 recession. The results of the present study show that increases in the following covariates yield protection against failure: company size as measured by market capitalization, debt-asset ratio, and price-to-book value ratio. In addition, the financial sector effect was found to be significant covariate influencing bankruptcy probability. Furthermore, this study compares the Cox proportional hazard model to a logistic model. The results of model fit and classification accuracy indicate that the former analysis outperforms the latter analysis.

KEY WORDS: Survival Analysis, Cox Regression, Business Failure Prediction

1. INTRODUCTION

Economic downturns expose businesses to serious stress and can lead to failure for some firms. Business failure prediction has been a salient topic in finance for both researchers and practitioners for decades. Using various statistical techniques, business failure prediction models attempt to estimate the bankruptcy probability of a firm using a set of covariates such as financial ratios and market-related variables (Beaver 1966; Altman 1968; Ohlson 1980; Zmijewski 1984; Whalen 1991; Laitinen and Luoma 1991; Shumway 2001; Chava and Jarrow 2004; Laitinen 2005; LeClere 2005; Jaggia and Thosar 2005; Gepp and Kumar 2008).

This study applies the Cox proportional hazards model (PH) to business failure cases during the recent US recession to examine the survival probability of a firm with a set of covariates. According to the Business Cycle Dating Committee of the National Bureau of Economic Research (NBER), a US recession began in December 2007. The firm specific covariates used in this study come from publicly available data just prior to the start of the study period. The sample consists of public companies that filed Chapter 11 bankruptcy, and a random selection of their surviving competitors, from the beginning of the first quarter of 2008 and running through the end of the first quarter of 2009.

Furthermore, this study compares the Cox PH model to an alternative statistical method, a binary logit model. This paper is organized as follows. In the second section, a brief review of the Cox PH model is presented. Section three describes the data and covariates used in the model. In the fourth section, model results are discussed in detail including the classification accuracy of both Cox PH and logistic models. Finally, a brief summary and conclusions are provided.

2. BACKGROUND – SURVIVAL ANALYSIS

Survival analysis consists of non-parametric analysis and a semi-parametric Cox proportional hazard model. The Kaplan-Meier, or product limit, method is used to estimate survival probabilities. The survival function, $S(t)$, represents the probability that a business will survive past a certain time *t* in financial distress models. For more details, see Hosmer and Lemeshow (1999) and Kleinbaum and Klein (2005).

The Cox proportional hazard (PH) model expresses the individual's hazard at time *t* with a set of covariates as shown in the following formula

$$h(t, X) = h_0(t) \exp\{\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_\rho X_\rho\}.$$

The probability of an individual's survival at time *t* is then calculated as follows

$$S(t, X) = \exp\left\{-\int_0^t h(u, X)du\right\}.$$

One prominent feature of the Cox PH model is its semi-parametric form. No assumptions are made about the baseline hazard function, $h_0(t)$, thus, a non-parametric part of the model. The parametric part of the model expresses the hazard rate as a function of the covariates. The coefficients, β_i, are estimated by maximizing a likelihood function. The Cox likelihood is called “partial likelihood function.” Rather than maximizing the full likelihood function, Cox's proposal was to maximize the function with respect to the parameters of interest only (1972).

The Cox PH model allows censored observations and incorporates survival times. A Cox PH model therefore uses more information than a logistic regression model could. Furthermore, an extended Cox PH model can be employed if the set of covariates are time-dependent (or time-varying). For more details, see references (Hosmer and Lemeshow 1999; Kleinbaum and Klein 2005).

.

3. DATA AND COVARIATES

The sample in the present study consists of public companies that filed Chapter 11 bankruptcy, and a random selection of their surviving competitors, from the beginning of the first quarter of 2008 and running through the end of the first quarter of 2009 (five quarters of daily bankruptcy data). Gathered data is from the fourth quarter of 2007 and are inferred to be a measure of company health near the official onset of the recession.

In this study, once a bankrupt firm was identified during the study period, the firm's Standard Industrial Classification (SIC) code was used to randomly select a non-bankrupt firm with the same industry code. The SIC codes, annual reports, and 10-K reports are available to the public via the EDGAR database maintained by the U.S. Securities and Exchange Commission (SEC). Only publicly traded U.S. companies on the New York, American, and NASDAQ stock exchanges were considered. Companies that either merged or did not fail are considered censored. In order to compute financial and market ratios, quarterly financial statements and historical prices were downloaded from Hoover's (a Dun & Bradstreet company), Standard & Poor's, Yahoo! Finance, and MSN Money. Additional data sources used for this study include: LexisNexis database, BankruptcyData.com, Chapter11Library.com, American Bankruptcy Institute, and the Federal Deposit Insurance Corporation (FDIC).

The preliminary set of data consisted of 35 failed companies and 35 non-failed companies selected as outlined above. The three firms that failed during the first quarter of the study period were removed. The rationale was to eliminate the possibility of including companies that may have been on the verge of failure near (or before) the beginning the study period. Additionally, the three outliers were removed after examining the Martingale and deviance residuals. As a result, the final data set consists of 33 failed companies (10 financial firms) and 31 non-failed companies (8 financial firms).

Table 1 summarizes the set of covariates included in the study. It consists of a combination of financial and market variables. In addition, a dichotomous variable (with the value 1 indicating a financial sector company and the value 0 indicating a non-financial sector company) is included. Sector definitions are based on Standard & Poor's GICS® methodology ("Global Industry Classification Standard" 2006).

.

Table 1. Covariates Used in the Model

Covariate	Measure
Current ratio (CACL)	Liquidity
Working capital to total assets ratio (WCTA)	Operating liquidity
Return on assets ratio (ROA)	Profitability
Earnings before interest and taxes to total assets ratio (EBITTA)	Profitability
Sales to total assets ratio (STA)	Efficiency
Total equity to total liabilities ratio (TETL)	Equity-to-debt
Total liabilities to total assets ratio (TLTA)	Debt-to-asset
Price-to-book value ratio (PBV)	Investment valuation
Price-to-earnings ratio (PE)	Investment valuation
Common logarithm of market capitalization (LMC)	Company size
Financial sector (FIN)	Financial sector effect

The Cox PH model was fitted to analyze the effects of financial and market variables on the survival of the company. The response variable in a hazard model is the time of non-failed firms and an indicator variable for censorship status. The values of survival time are considered censored if the value of censorship status is 0; otherwise, they are considered failure times. The SAS® statistical software system was used extensively to estimate regression model parameters (SAS Institute Inc. 2004; Kim 2009).

4. SURVIVAL ANALYSIS MODEL RESULTS

4.1 Model Fit

Table 2 presents the results of estimated Cox PH regression model chosen by the stepwise selection process. The stepwise method identified company size (LMC), price-to-book value ratio (PBV), debt-asset ratio (TLTA), and a financial sector indicator variable (FIN) as significant covariates. The p-values of the parameter estimates for the regression coefficients are highly significant for the four covariates ($p < 0.0001$, $p = 0.0025$, $p = 0.0020$, and $p = 0.0001$ for LMC, PBV, TLTA and FIN, respectively).

To assess the overall goodness-of-fit, the log likelihood test is used. The likelihood-ratio test statistic equals 47.8246 with 4 degrees of freedom. The null hypothesis, that all covariates included in the model have zero coefficients, is rejected ($p < 0.0001$). Thus, it indicates that the result is highly significant. Another measure of model fit may be “some measure analogous to R^2” (Hosmer and Lemeshow 1999). In linear regression, the R^2 measure explains the proportion of variability of the response variable by the explanatory variables. The R^2 in the Cox regression is a pseudo measure of association between the response variable and covariates. The measure of R^2 in the present study is approximately equal to 0.7756. Overall, the model seems to provide a fairly good fit.

In order to assess the proportional hazard assumption, a statistical test is carried out by finding the correlation between the Schoenfeld residuals for a given covariate and the ranked individual failure time. Results indicate that there is no relation between the residuals and survival time, thus all covariates meet the PH assumption (p-values for the residuals of LMC, PBV, TLTA, and FIN are 0.7438, 0.2373, 0.4330, and 0.2748, respectively).

Table 2. Cox Proportional Hazard Regression Model

Model Fit Statistics								
		Criterion	Without Covariates	With Covariates				
		-2 LOG L	252.158	204.334				
		AIC	252.158	212.334				
		SBC	252.158	218.32				
Testing Global Null Hypothesis: BETA=0								
		Test	Chi-Square	DF	Pr > ChiSq			
		Likelihood Ratio	47.8246	4	<.0001			
		Score	49.1767	4	<.0001			
		Wald	39.7491	4	<.0001			
Analysis of Maximum Likelihood Estimates								
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
FIN	1	1.70764	0.44737	14.5702	0.0001	5.516	2.295	13.256
LMC	1	-1.25214	0.21731	33.2016	<.0001	0.286	0.187	0.438
TLTA	1	-0.92919	0.30043	9.5657	0.002	0.395	0.219	0.712
PBV	1	-0.04993	0.0165	9.1623	0.0025	0.951	0.921	0.983

4.2 Model Interpretation

The effect of the financial sector covariate was found to be significant. The interpretation of the estimated hazard ratio of 5.516 is that financial companies in this study (FIN = 1) fail at about five and half times the rate of those in non-financial sector. The 95% confidence interval for the hazard ratio suggests a factor as low as 2.3 or as high as 13.3

For the three continuous scale covariates (LMC, PBV, and TLTA), negative parameter estimates and hazard ratios of less than 1.0 indicate the association with a decrease in the risk to fail. Below, each significant covariate and its associated hazard ratio are discussed.

A company's size is often determined by its market capitalization. For example, market capitalization is used in asset allocation for stock mutual funds, where the size may be categorized as large, medium, or small cap. The data set used in this study has a wide range of market capitalizations. In such cases the common logarithm of market capitalization (LMC) is often a more useful quantity. The Cox PH model estimates that an increase of one unit in the logarithm of market capitalization shrinks the hazard rate by 71.4% (1 - 0.286 = 0.714). That is, if all other covariates are fixed, a company that is 10 times larger in terms of market capitalization will have a 71.4% reduced risk of failing. The 95% confidence interval of the hazard ratio is given by the range of values 0.187 - 0.438. Thus, the company that is 10 times larger in terms of market capitalization, with all other covariates fixed, will have a 56.2% to 81.3% reduced risk of failing.

The debt-asset ratio (TLTA) is obtained by dividing the total liabilities by total assets. This ratio provides the proportion of the firm's assets that are financed through debt. Thus, a high debt-asset ratio is associated with a highly leveraged firm. Although there are some benefits of financing through debt over equity, debt financing is recommended only up to a certain level. The Cox PH model shows that the debt-asset ratio provides a protective measure against the risk of failing. A unit increase in TLTA is unlikely, but an increase of 10%, with all other covariates fixed, will reduce the risk of failing by nearly 8.9% ($1 – e^{0.1 (-0.92919)} = 0.0887$). For a 10% increase in TLTA, the 95% confidence interval of the hazard ratio is given by the range of values 0.859 - 0.967. Therefore, a 10% increase in TLTA, with all other covariates fixed, will have a 3.3% to 14.1% reduced risk of failure.

A price-to-book value (PBV) ratio is used to compare a stock's market price to a company's book value per share. The Cox PH model indicates that an increase of one unit in the PBV ratio, with all other covariates fixed, will reduce the risk of failing by 4.9% (1 - 0.951 = 0.049). The 95% confidence interval of the hazard ratio is given by the range of values 0.921 - 0.983. Therefore, a unit increase in PBV, with all other covariates fixed, will have a 1.7% to 7.9% reduced risk of failure.

4.3 Model Accuracy

Several studies about model classification accuracy can be found in the literature (Beaver 1966; Altman1968; Ohlson1980; Laitinen and Luoma1991; Laitinen 2005). Ideally, model accuracy could be assessed by using a holdout sample of data. Unfortunately, the very nature of this study (small data set) precludes the use of a holdout sample. Therefore, the entire data set is used to assess the accuracy of the model.

For this study, prediction accuracy was evaluated by first arranging Cox PH model estimated survival probabilities into quartiles based on rank. Next, I assume that the upper two quartiles contain the firms that the model indicates are most likely to survive, whereas the lower two quartiles contain the firms that are least likely to survive. This analysis is similar that followed by Laitinen and Luoma (1991). Table 3 below shows the results of this classification. The overall classification accuracy when measured in this manner is 82.81%.

Table 3. Cox PH Regression Model Classification Accuracy

Cox PH Regression Model		
Most Likely to Survive	Upper 50%	% Correctly Classified as Not Failed 81.25%
Most Likely to Fail	Lower 50%	% Correctly Classified as Failed 84.38%

Looking at the classification accuracy at each quartile yields the results shown in Table 4. The upper and two lower quartiles yield a combined classification accuracy of 89.6% for the given subset of firms.

Table 4. Cox PH Regression Model Classification Accuracy by Quartile

	Quartile	Cox PH Regression Model
Most Likely to Survive	Q1	% Correctly Classified as Not Failed 100.00%
Likely to Survive	Q2	% Correctly Classified as Not Failed 62.50%
Likely to Fail	Q3	% Correctly Classified as Failed 81.30%
Most Likely to Fail	Q4	% Correctly Classified as Failed 87.50%

4.4 Model Use

The Cox PH model accuracy results presented in Table 4 motivated an investigation into using a confidence interval approach for failure prediction. Figure 1 presents the Cox PH model estimated survival curve, and its 95% confidence interval, using the median covariate values of failed companies in the study.

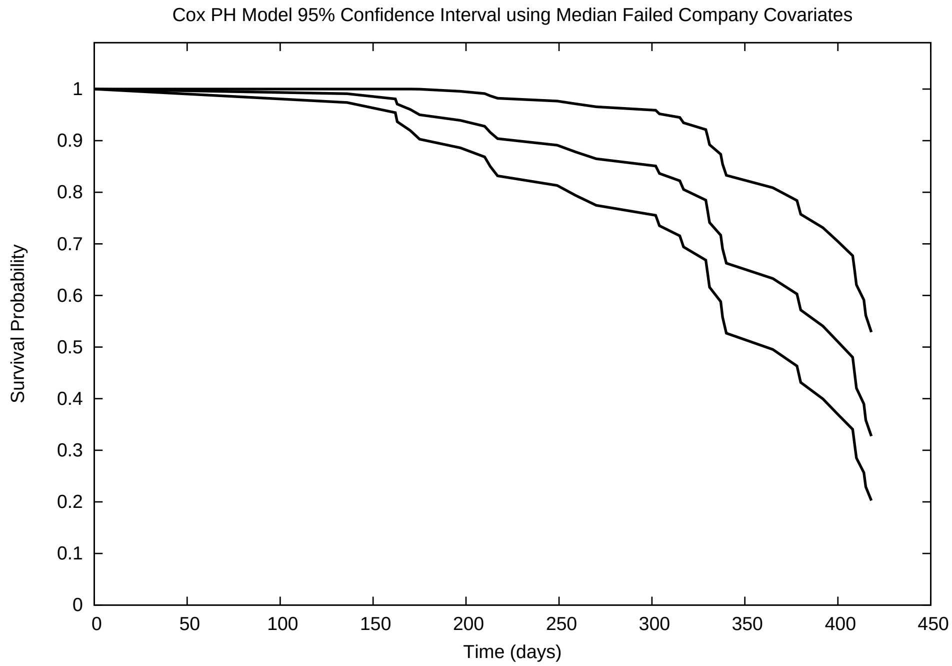

Figure 1. Survival Curve and 95% Confidence Interval using Median Failed Company Covariates

Figure 2 shows the failed companies in this study plotted along with the Cox PH 95% survival curve interval calculated using the median covariates of failed companies. Note that all but one of the failed companies fall below, or on, the upper limit of the confidence interval.

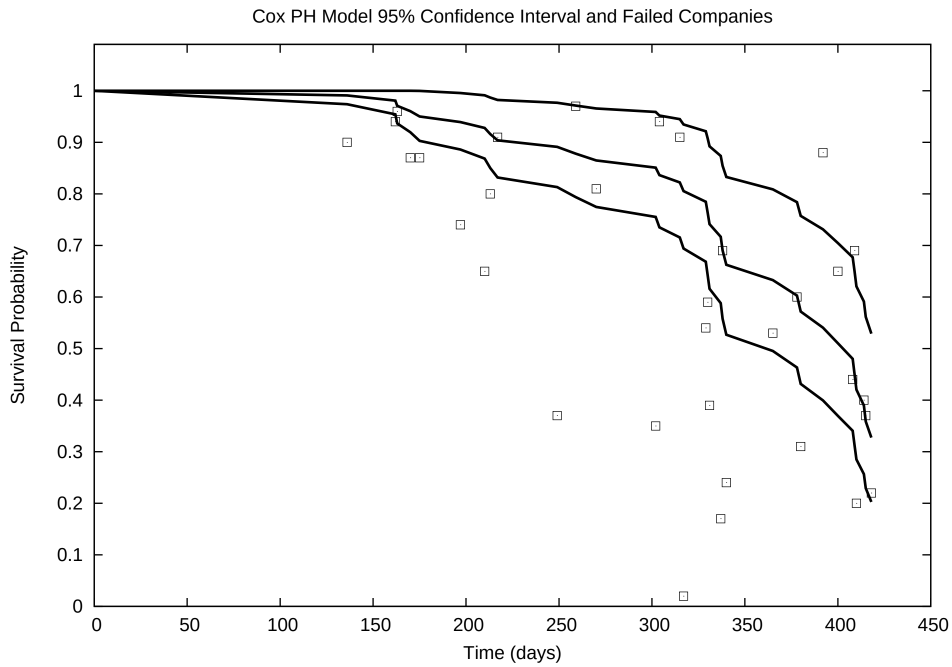

Figure 2. 95% Survival Curve Confidence Interval and Failed Companies

Finally, Figure 3 shows the Cox PH model predicted survival curves for failed companies in this study plotted along with the 95% survival curve interval calculated using the median covariates of failed companies. Notice that the predicted survival curves for all of the failed companies fall under the upper limit of the confidence interval at some time during the study period.

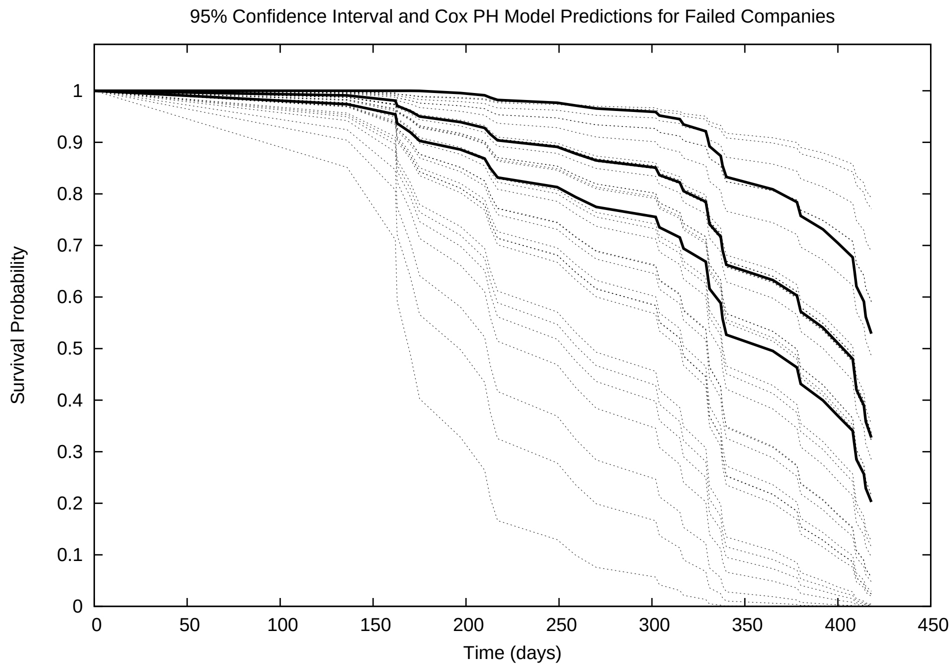

Figure 3. 95% Survival Curve Confidence Interval and Cox PH Predicted Survival Curves for Failed Companies

Taken together, Figure 2 and Figure 3 suggest that the central value of the confidence interval may be a useful indicator of survival (above) or failure (below) – at least for the data in this study. Table 5 below shows the results of this classification. The overall classification accuracy when measured in this manner is 87.50%. This classification accuracy is larger than 82.81% observed using the method in Table 3.

Table 5. Cox PH Regression Model Alternative Classification

Cox PH Regression Model		
Most Likely to Survive	Above Median Survival Curve	% Correctly Classified as Not Failed 82.86%
Most Likely to Fail	Below Median Survival Curve	% Correctly Classified as Failed 93.10%

5. COMPARISON TO A LOGISTIC MODEL

Logistic regression models are often used in modeling financial distress. The general form of a logistic regression model, where $\pi(X)$ is the probability of not failing, is

$$\log\left\{\frac{\pi(X)}{1-\pi(X)}\right\} = \beta_0 + \beta_1 X_1 + \cdots + \beta_\rho X_\rho \,.$$

Below, a logistic model is constructed and compared to the Cox PH model already presented.

5.1 Model Fit

Table 6 shows the estimated parameters of the logistic model chosen by the stepwise selection process. Unlike the survival analysis model presented earlier, a model based on logistic regression selects only company size (LMC) as a significant covariate influencing the survival of a company.

The Chi-square test statistic shows that the overall logistic regression model is highly significant. The likelihood ratio statistic is 21.0581 with 1 degree of freedom. Thus, the null hypothesis in favor of the model with intercept only is rejected ($p < 0.0001$). The measure of R^2 and max-rescaled R^2 are 0.2804 and 0.3740, respectively. In comparison, the Cox PH model has a much better fit as indicated by its higher measure of R^2 (0.7756 versus 0.3740).

Logistic regression results indicate that the p-values of the parameter estimates for the regression coefficients are highly significant for both intercept and LMC covariate ($p = 0.0003$ and $p = 0.0002$). A positive parameter estimate implies the association with an increase in survival probability.

Table 6. Logistic Regression Model

		Model Fit Statistics			
	Criterion	Intercept Only	Intercept and Covariates		
	AIC	90.66	71.602		
	SC	92.819	75.92		
	-2 Log L	88.66	67.602		
		Testing Global Null Hypothesis: BETA=0			
	Test	Chi-Square	DF	Pr > ChiSq	
	Likelihood Ratio	21.0581	1	<.0001	
	Score	18.1768	1	<.0001	
	Wald	13.5561	1	0.0002	
R-Square	0.2804		Max-rescaled R-Square	0.374	
		Hosmer and Lemeshow Goodness-of-Fit Test			
	Chi-Square	DF	Pr > ChiSq		
	14.6289	9	0.1016		
		Analysis of Maximum Likelihood Estimates			
Parameter	Estimate	Standard Error	Wald Chi-Square	DF	Pr > ChiSq
Intercept	-3.1337	0.8764	12.785	1	0.0003
LMC	1.1003	0.2988	13.5561	1	0.0002
		Odds Ratio Estimates			
	Effect	Point Estimate	95% Wald Confidence Limits		
	LMC	3.005	1.673	5.398	

5.2 Model Interpretation

The Logistic model estimates that an increase of one unit in the logarithm of market capitalization (LMC) shrinks the hazard rate by 66.7% (1 – 1/3.005 = 0.667). This implies that a company that is 10 times larger in terms of market capitalization will have a 66.7% reduced risk of failing. The 95% confidence interval of the hazard ratio is given by the range of values 0.185 – 0.598 (1/5.398, 1/1.673). Thus, the company that is 10 times larger in terms of market capitalization will have a 40.2% to 81.5% reduced risk of failing. In comparison, the Cox PH model indicated that an increase of one unit in LMC shrinks the hazard rate by 71.4%. The Cox PH 95% confidence interval is 56.2% to 81.3% reduced risk of failing.

5.3 Model Accuracy

Model accuracy of the logistic model is compared to Cox PH model in Table 7 below. The Cox PH model outperforms the logistic model in this comparison. The overall accuracy for each model compared in this manner is 82.81% (Cox PH) and 73.44% (Logistic). The additional significant covariates in the Cox PH model are largely responsible for its superior model accuracy.

Table 7. Model Accuracy Comparison

		Cox PH Regression Model	Logistic Regression Model
Most Likely to Survive	Upper 50%	% Correctly Classified as Not Failed 81.25%	% Correctly Classified as Not Failed 71.88%
Most Likely to Fail	Lower 50%	% Correctly Classified as Failed 84.38%	% Correctly Classified as Failed 75.00%

Figure 4 shows the failed companies in this study plotted along with the estimated logistic survival probability and 95% confidence interval using the median covariate value of failed companies. Notice that, unlike the Cox PH model, a significant number of failed companies lie above the confidence interval.

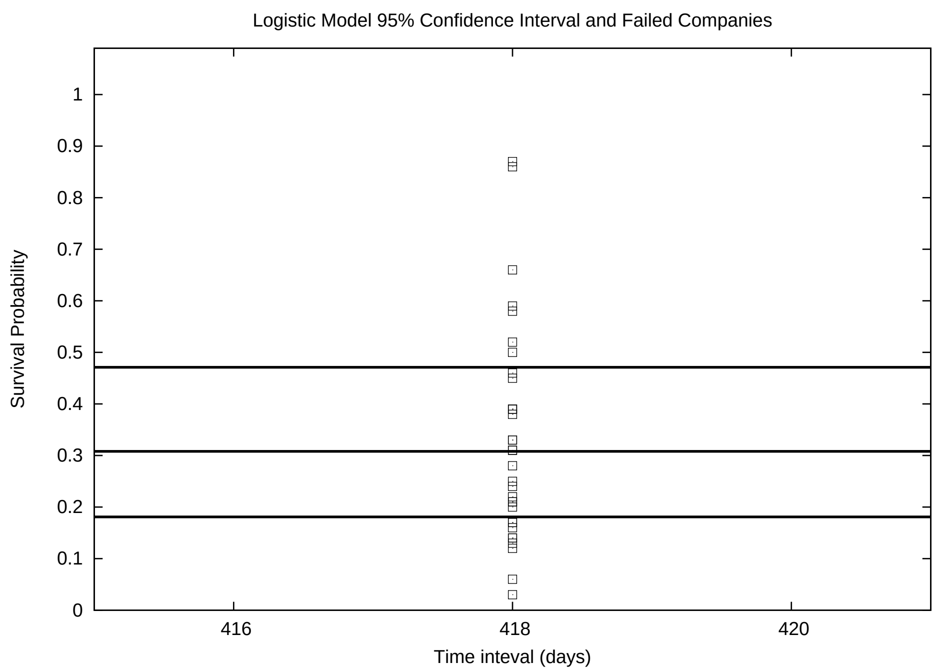

Figure 4. Logistic Model 95% Confidence Interval and Failed Companies

6. SUMMARY AND CONCLUSIONS

This study attempted to model company failure during the first five quarters of the recession that began in late 2007. A survival analysis approach using a Cox PH model was successfully used and found to outperform a logistic model in terms of model fit and classification of companies into failed or not-failed categories.

There are methodological differences between the two techniques. The Cox PH model accounts for survival times and therefore intrinsically uses more information compared to the logistic model. The latter analysis is cross-sectional, whereas the former is longitudinal.

The Cox PH model included significant market and financial covariates while the logistic model selected only one market covariate and no financial covariates. The Cox PH model indicated that financial firms were far more likely to fail during the study period and found that increases in company size (as measured by market capitalization), debt-asset ratio (TLTA), and price-to-book value (PBV) were protective factors against failure. Furthermore, the Cox PH model's 95% confidence interval calculated using the median covariate values from failed companies may be a useful tool in failure prediction.

REFERENCES

Altman, E. I. (1968), "Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy," *Journal of Finance*, 23, 589-609.

Beaver, W. H. (1966), "Financial Ratios as Predictors of Failure. Empirical Research in Accounting: Selected Studies," *Journal of Accounting Research*, 4(Supplement), 71-127.

Chava, S., and Jarrow, R. A. (2004) "Bankruptcy Prediction with Industry Effects," *Review of Finance*, *8* (4), 537-569.

Cox, D. R. (1972), “Regression Models and Life Tables (with discussion),” *Journal of Royal Statistical Society*, Series B (Methodological), 34, 187-220.

Gepp, A., and Kumar, K. (2008), “The Role of Survival Analysis in Financial Distress Prediction,” *International Research Journal of Finance and Economics*, 16, 1450-2887.

Hosmer JR., D. W., and Lemeshow, S. (1999), *Applied Survival Analysis: Regression Modeling of Time to Event Data*, New York: John Wiley & Sons.

Jaggia, S., and Thosar, S. (2005), “Survival Analysis with Artificially Constructed Events,” *Review of Accounting and Finance*, 4 (4), 34-49.

Kim, O. (2009), “Developing Business Failure Prediction Models Using SAS® Software,” Statistics and Data Analysis Section of SouthEast SAS Users Group 2009 Conference.

Kleinbaum, D. G., and Klein, M. (2005), *Survival Analysis: A Self-Learning Text* (2nd ed.), New York: Springer.

Laitinen, E. K., and Luoma, M. (1991), “Survival Analysis as a Tool for Company Failure Prediction,” *Omega, The International Journal of Management Science*, 19 (6), 673-678.

Laitinen, E. K. (2005), “Survival Analysis and Financial Distress Prediction: Finish Evidence,” *Review of Accounting & Finance*, 4 (4), 76-90.

Ohlson, J. A. (1980), “Financial Ratios and the Probabilistic Prediction of Bankruptcy,” *Journal of Accounting Research*, 18 (1), 109-131.

LeClere, M. J. (2005), “Time-Dependent and Time-Invariant Covariates within a Proportional Hazards

Model: A Financial Distress Application," *Review of Accounting & Finance*, 4 (4), 91.

SAS Institute Inc. (2004), *SAS/STAT ® 9.1 User's Guide*, Volumes 1-7, Cary, NC: SAS Institute Inc..

Shumway, T. (2001), "Forecasting Bankruptcy More Accurately: A Simple Hazard Model," *The Journal of Business*, 74 (1), 101-124.

Standard & Poor's (2006), "Global Industry Classification Standard (GICS®)." Available at http://www2.standardandpoors.com/spf/pdf/index/GICS_methodology.pdf.

Whalen, G. A (1991), "Proportional Hazards Model of Bank Failure: An Examination of Its Usefulness as an Early Warning Tool," *Federal Reserve Bank of Cleveland, Economic Review*, Quarter 1, 21-31.

Zmijewski, M.E (1984), "Methodological Issues Related to the Estimation of Financial Distress Prediction Models," *Journal of Accounting Research*, 22, 59-82.