Recommendation

This report recommends that a Cox Proportional-Hazards survival analysis model be used in predicting the likelihood of a company filing for bankruptcy in the next 16 years. Further, the variables that were considered to be significant and impactful in the model are net debt to EBITDA ratio, current ratio, debt to cash flow from operations, and market capitalization in the most recent year (in this case 2006).

Methodology

The first step consisted of cleaning the data and removing rows which contained missing information. Second, preliminary review of the survival data was performed by plotting the log probability of survival beyond each time step against time. Third, the appropriate survival analysis model was selected and variables which were not significant or had a hazard ratio of one were reviewed before removal. Finally, as a comparison, alternative survival analysis models were also computed.

Preliminary Analysis

The survival plot is provided in Exhibit 1. The survival plot is somewhat linear but does contain varying slopes indicating a changing lambda overtime. Had the scatterplot indicated the relationship were linear then an Exponential distribution would have been appropriate. A changing lambda initially indicated a Weibull distribution would be appropriate, however the slopes seem to both increase and decrease overtime rather than only increasing or decreasing. Therefore, the most appropriate distribution is the Cox Proportional-Hazard model. This model allows for fluctuation of lambda overtime.

The main drawback of the Cox Proportional-Hazard Model is that it does not allow for predicting probability of survival past a particular timestep. However, the Cox Proportional-Hazard Model enables estimation of hazard ratios, which provides insights into variables which influences survival. Therefore, a Weibull distribution model will also be provided to calculate probabilities of survival past particular timesteps. In this case, the hazard ratios provided by the Weibull distribution model is likely less accurate than those provided by the Cox Proportional-Hazard model.

Cox Proportional-Hazard Analysis

The Cox Proportional-Hazard output is provided in Exhibit 2. Variables were removed which were insignificant or did not impact the survival of the company. No impact to survival is indicated by a hazard ratio of one. The variables which remained are net debt to EBITDA ratio, debt to cash flow from operations, and market capitalization of most recent year. There were many variables which were removed due to lack of significance and likelihood of providing the same information as other variables which were already in the model.

An example of a removed variable is the market capitalization of the second most recent year. This information is likely already provided in the most recent year and out of date information is less important since companies may change significantly in a year due to external factors. However, the most recent market capitalization is an important factor to include in the model since companies which are less likely to file for bankruptcy typically have a higher market capitalization due to efficient markets.

Other factors which include various forms of earnings calculations were also removed such as gross profit and net income. The most important factors to a company filing for bankruptcy is the ability to fulfill current financial obligations based on earnings rather than only earnings. This information is captured by ratios in the model such as net debt to EBITDA, and debt to cash flow from operations.

Investigation of the hazard ratios show that a high net debt to EBITDA ratio negatively impacts survival. This is logical since it shows you likely have too much debt to repay based on earnings. This is in contrast to a high debt to cash flow from operations and high market capitalization which positively impacts survival.

Weibull Distribution Analysis

The Weibull distribution model is provided in Exhibit 3. In this model, variables which did not have hazard ratios equal to one remained. The variables which remained were net debt to EBITDA ratio, profit margin, and current ratio. Similar to the Cox Proportional-Hazard model, net debt to EBITDA ratio and the current ratio captures the ability of a company to meet immediate financial obligations. Further, in terms of profit margin, a high value indicates likelihood of future success.

Similar to the Cox Proportional-Hazard model, investigation of the hazard ratios show that a high net debt to EBITDA ratio negatively impacts survival. This is in contrast to profit margin and current ratio which positively impacts survival.

Exponential Distribution Analysis

For completeness, an Exponential distribution model is provided in Exhibit 4. Similar to the previous analyses, variables determined to be insignificant or explaining information similar to other variables which already exist in the model were removed. As mentioned previously, a Weibull distribution is most appropriate for survival probability calculations.

Big Data Analytics – Assignment 4
Student ID: 250830704

Exhibit 1: Survival Plot (K-M Model)

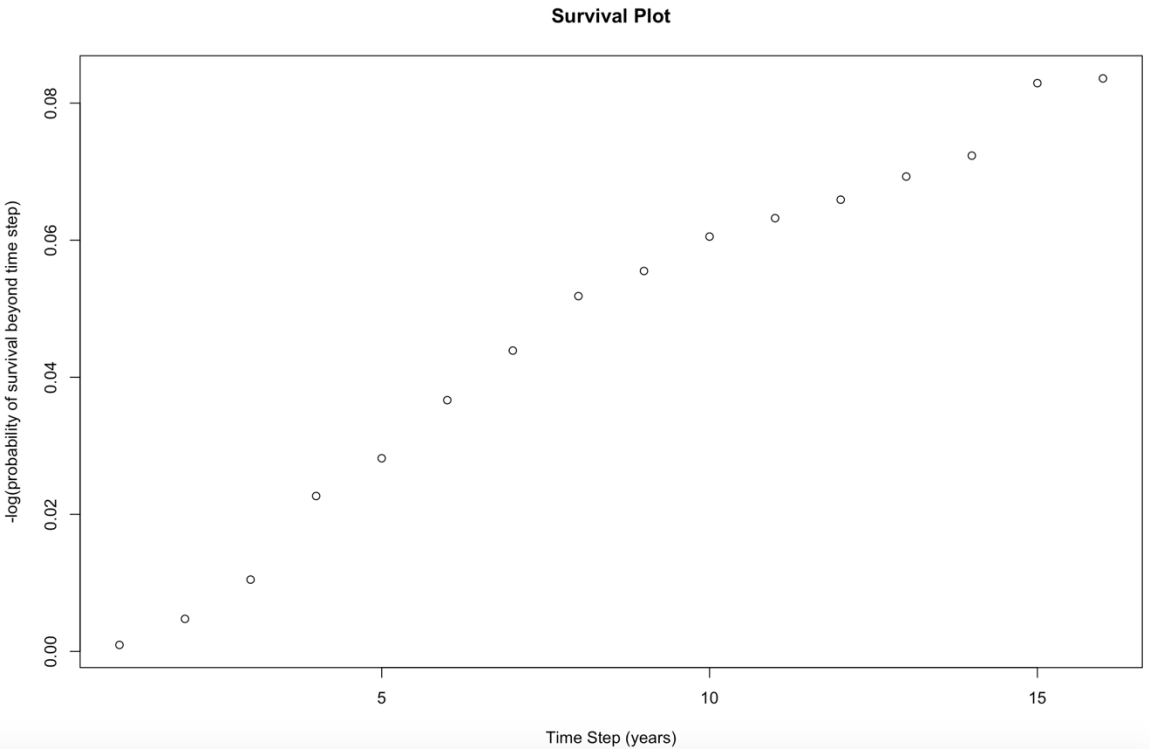

Exhibit 2: Cox Proportional-Hazard Output

```
Call:
coxph(formula = Surv(Years_to_Bankrupt, Bankruptcy) ~ Net_Debt.EBITDA +
    Debt_to_Cash_Flow_From_Ops + Market_Capitalization.1, data = bankruptcy.df)

  n= 3167, number of events= 254

                                 coef  exp(coef)   se(coef)      z Pr(>|z|)
Net_Debt.EBITDA             3.157e-02  1.032e+00  3.563e-03  8.861  < 2e-16 ***
Debt_to_Cash_Flow_From_Ops -7.746e-04  9.992e-01  1.102e-04 -7.032 2.04e-12 ***
Market_Capitalization.1    -9.698e-05  9.999e-01  2.659e-05 -3.648 0.000265 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

                           exp(coef) exp(-coef) lower .95 upper .95
Net_Debt.EBITDA               1.0321     0.9689    1.0249    1.0393
Debt_to_Cash_Flow_From_Ops    0.9992     1.0008    0.9990    0.9994
Market_Capitalization.1       0.9999     1.0001    0.9999    1.0000

Concordance= 0.659  (se = 0.017 )
Likelihood ratio test= 87.94  on 3 df,   p=<2e-16
Wald test            = 152.3  on 3 df,   p=<2e-16
Score (logrank) test = 360.8  on 3 df,   p=<2e-16
```

Big Data Analytics – Assignment 4
Student ID: 250830704

Exhibit 3: Weibull Distribution Model

```
> ConvertWeibull(weibull.out)
$vars
                    Estimate           SE
lambda           0.003097928 0.0006785087
gamma            1.191348118 0.0732296600
Net_Debt.EBITDA  0.030495656 0.0036189685
Profit_Margin   -0.840214591 0.3126763271
Current_Ratio   -0.016055295 0.0208683573

$HR
                        HR        LB        UB
Net_Debt.EBITDA 1.0309654 1.0236786 1.0383041
Profit_Margin   0.4316179 0.2338562 0.7966178
Current_Ratio   0.9840729 0.9446351 1.0251572

$ETR
                       ETR        LB       UB
Net_Debt.EBITDA 0.9747272 0.9683140 0.981183
Profit_Margin   2.0243805 1.2019776 3.409478
Current_Ratio   1.0135678 0.9793303 1.049002

>
```

Exhibit 4: Exponential Distribution Model

```
Call:
survreg(formula = Surv(Years_to_Bankrupt, Bankruptcy) ~ Total_Revenue +
    Gross_Profit + Net_Debt.EBITDA + Net_Debt + Current_Ratio +
    Debt_to_Cash_Flow_From_Ops + Market_Capitalization.1, data = bankruptcy.df,
    dist = "exponential")
                              Value Std. Error     z       p
(Intercept)                5.08e+00   9.38e-02 54.15 < 2e-16
Total_Revenue              5.48e-05   2.75e-05  1.99   0.046
Gross_Profit              -5.10e-04   1.10e-04 -4.63 3.7e-06
Net_Debt.EBITDA           -2.95e-02   3.71e-03 -7.96 1.7e-15
Net_Debt                  -4.18e-05   1.81e-05 -2.31   0.021
Current_Ratio              2.56e-02   2.28e-02  1.13   0.260
Debt_to_Cash_Flow_From_Ops 5.65e-04   1.02e-04  5.56 2.6e-08
Market_Capitalization.1    2.35e-04   4.45e-05  5.28 1.3e-07

Scale fixed at 1

Exponential distribution
Loglik(model)= -1534.4   Loglik(intercept only)= -1588.6
        Chisq= 108.45 on 7 degrees of freedom, p= 1.9e-20
Number of Newton-Raphson Iterations: 9
n= 3167
```

Big Data Analytics – Assignment 4
Student ID: 250830704

Exhibit 5: R Code

```
library(survival)
library(SurvRegCensCov)
bankruptcy.df <- read.csv("/Users/chriskwan/Documents/R/RLabs/Bankruptcy Data.csv")
bankruptcy.df <- na.omit(bankruptcy.df)
#NA row checker
bankruptcy.df.na <- bankruptcy.df[rowSums(is.na(bankruptcy.df))>0,]
bankruptcy.df.na
#Plot survival first for initial diagnosis
out <- survfit(Surv(Years_to_Bankrupt,Bankruptcy)~1,data=bankruptcy.df)
y<-out$surv
y<- -log(y)
t<- out$time
plot(t,y,xlab = "Time Step (years)", ylab = "-log(probability of survival beyond time step)", main = "Survival Plot")

#Exponential
exp.out <- survreg(Surv(Years_to_Bankrupt,Bankruptcy)~Total_Revenue+Gross_Profit+Net_Income+EBITDA+Total_Current_Assets+Total_Assets+Total_Current_Liabilities+Cash_from_Ops.+Net_Debt.EBITDA+Profit_Margin+Net_Debt+Net_Worknq_Capital+Current_Ratio+Debt_to_Cash_Flow_From_Ops+Market_Capitalization+Market_Capitalization.1, data=bankruptcy.df , dist='exponential')
summary(exp.out)
#Some of these variables tell us the same thing, remove EBITDA since we have it included in Net_DebtEBITDA and it is similar to Net Income
exp.out <- survreg(Surv(Years_to_Bankrupt,Bankruptcy)~Total_Revenue+Gross_Profit+Net_Income+Total_Current_Assets+Total_Assets+Total_Current_Liabilities+Cash_from_Ops.+Net_Debt.EBITDA+Profit_Margin+Net_Debt+Net_Worknq_Capital+Current_Ratio+Debt_to_Cash_Flow_From_Ops+Market_Capitalization+Market_Capitalization.1, data=bankruptcy.df , dist='exponential')
summary(exp.out)
#Current ratio is current assets/current liabilities. Therefore we can remove both as they would indicate the same thing
exp.out <- survreg(Surv(Years_to_Bankrupt,Bankruptcy)~Total_Revenue+Gross_Profit+Net_Income+Total_Assets+Cash_from_Ops.+Net_Debt.EBITDA+Profit_Margin+Net_Debt+Net_Worknq_Capital+Current_Ratio+Debt_to_Cash_Flow_From_Ops+Market_Capitalization+Market_Capitalization.1, data=bankruptcy.df , dist='exponential')
summary(exp.out)
#Net working capital generally tells the same thing as current ratio so lets remove it
exp.out <- survreg(Surv(Years_to_Bankrupt,Bankruptcy)~Total_Revenue+Gross_Profit+Net_Income+Total_Assets+Cash_from_Ops.+Net_Debt.EBITDA+Profit_Margin+Net_Debt+Current_Ratio+Debt_to_Cash_Flow_From_Ops+Market_Capitalization+Market_Capitalization.1, data=bankruptcy.df , dist='exponential')
summary(exp.out)
#Market Cap for a company that is might be bankrupt in previous year is more relevant than 2 years ago and likely says something similar. The market is dynamic and things can change quickly in a company. The current market cap probably says enough about the company.
exp.out <- survreg(Surv(Years_to_Bankrupt,Bankruptcy)~Total_Revenue+Gross_Profit+Net_Income+Total_Assets+Cash_from_Ops.+Net_Debt.EBITDA+Profit_Margin+Net_Debt+Current_Ratio+Debt_to_Cash_Flow_From_Ops+Market_Capitalization.1, data=bankruptcy.df , dist='exponential')
summary(exp.out)
#When a company goes bankrupt total assets generall doesn't matter it's cash that matters most. If you have a lot of assets you can still go bankrupt, it's just that you might be more likely to pay creditors with assets you have, so lets remove it
exp.out <- survreg(Surv(Years_to_Bankrupt,Bankruptcy)~Total_Revenue+Gross_Profit+Net_Income+Cash_from_Ops.+Net_Debt.EBITDA+Profit_Margin+Net_Debt+Current_Ratio+Debt_to_Cash_Flow_From_Ops+Market_Capitalization.1, data=bankruptcy.df , dist='exponential')
summary(exp.out)
#Net income and cash from operations are closely intertwind and a high net income usually means more cash from operations. Both say similar things to Gross Profit which is more significant. Profit Margin is also essentially Gross Profit. If a company is going bankrupt Gross Profit is more important since they can always slash costs to stay afloat
exp.out <- survreg(Surv(Years_to_Bankrupt,Bankruptcy)~Total_Revenue+Gross_Profit+Net_Debt.EBITDA+Net_Debt+Current_Ratio+Debt_to_Cash_Flow_From_Ops+Market_Capitalization.1, data=bankruptcy.df , dist='exponential')
summary(exp.out)

#Try Weibull
weibull.out <- survreg(Surv(Years_to_Bankrupt,Bankruptcy)~Total_Revenue+Gross_Profit+Net_Income+EBITDA+Total_Current_Assets+Total_Assets+Total_Current_Liabilities+Cash_from_Ops.+Net_Debt.EBITDA+Profit_Margin+Net_Debt+Net_Worknq_Capital+Current_Ratio+Debt_to_Cash_Flow_From_Ops+Market_Capitalization+Market_Capitalization.1, data=bankruptcy.df , dist='weibull')
weibull.out
ConvertWeibull(weibull.out)

#From here the hazard ratios which are not 1 are Net_Debt.EBITDA, Profit Margin, and Current Ratio
weibull.out <- survreg(Surv(Years_to_Bankrupt,Bankruptcy)~Net_Debt.EBITDA+Profit_Margin+Current_Ratio, data=bankruptcy.df , dist='weibull')
weibull.out
ConvertWeibull(weibull.out)

#Cox proportional Hazard Model
cox.out <- coxph(Surv(Years_to_Bankrupt,Bankruptcy)~Total_Revenue+Gross_Profit+Net_Income+EBITDA+Total_Current_Assets+Total_Assets+Total_Current_Liabilities+Cash_from_Ops.+Net_Debt.EBITDA+Profit_Margin+Net_Debt+Net_Worknq_Capital+Current_Ratio+Debt_to_Cash_Flow_From_Ops+Market_Capitalization+Market_Capitalization.1, data=bankruptcy.df)
summary(cox.out)
#Remove EBITDA
cox.out <- coxph(Surv(Years_to_Bankrupt,Bankruptcy)~Total_Revenue+Gross_Profit+Net_Income+Total_Current_Assets+Total_Assets+Total_Current_Liabilities+Cash_from_Ops.+Net_Debt.EBITDA+Profit_Margin+Net_Debt+Net_Worknq_Capital+Current_Ratio+Debt_to_Cash_Flow_From_Ops+Market_Capitalization+Market_Capitalization.1, data=bankruptcy.df)
summary(cox.out)
```

Big Data Analytics – Assignment 4
Student ID: 250830704

```
#Remove total current assets and liabilities
cox.out <- coxph(Surv(Years_to_Bankrupt,Bankruptcy)~Total_Revenue+Gross_Profit+Net_Income+Total_Assets+Cash_from_Ops. +
Net_Debt.EBITDA+Profit_Margin+Net_Debt+Net_Workng_Capital+Current_Ratio+Debt_to_Cash_Flow_From_Ops+Market_Capitalization+Market_Capitalization.
1, data=bankruptcy.df)
summary(cox.out)
#Remove cash from Ops
cox.out <- coxph(Surv(Years_to_Bankrupt,Bankruptcy)~Total_Revenue+Gross_Profit+Net_Income+Total_Assets+
Net_Debt.EBITDA+Profit_Margin+Net_Debt+Net_Workng_Capital+Current_Ratio+Debt_to_Cash_Flow_From_Ops+Market_Capitalization+Market_Capitalization.
1, data=bankruptcy.df)
summary(cox.out)
#Remove Total Assets
cox.out <- coxph(Surv(Years_to_Bankrupt,Bankruptcy)~Total_Revenue+Gross_Profit+Net_Income+
Net_Debt.EBITDA+Profit_Margin+Net_Debt+Net_Workng_Capital+Current_Ratio+Debt_to_Cash_Flow_From_Ops+Market_Capitalization+Market_Capitalization.
1, data=bankruptcy.df)
summary(cox.out)
#Remove Net Working Capital
cox.out <- coxph(Surv(Years_to_Bankrupt,Bankruptcy)~Total_Revenue+Gross_Profit+Net_Income+
Net_Debt.EBITDA+Profit_Margin+Net_Debt+Current_Ratio+Debt_to_Cash_Flow_From_Ops+Market_Capitalization+Market_Capitalization.1, data=bankruptcy.df)
summary(cox.out)
#Remove Net Income
cox.out <- coxph(Surv(Years_to_Bankrupt,Bankruptcy)~Total_Revenue+Gross_Profit+
Net_Debt.EBITDA+Profit_Margin+Net_Debt+Current_Ratio+Debt_to_Cash_Flow_From_Ops+Market_Capitalization+Market_Capitalization.1, data=bankruptcy.df)
summary(cox.out)
#Remove Market Cap
cox.out <- coxph(Surv(Years_to_Bankrupt,Bankruptcy)~Total_Revenue+Gross_Profit+
Net_Debt.EBITDA+Profit_Margin+Net_Debt+Current_Ratio+Debt_to_Cash_Flow_From_Ops+Market_Capitalization.1, data=bankruptcy.df)
summary(cox.out)
#Remove Profit Margin
cox.out <- coxph(Surv(Years_to_Bankrupt,Bankruptcy)~Total_Revenue+Gross_Profit+
Net_Debt.EBITDA+Net_Debt+Current_Ratio+Debt_to_Cash_Flow_From_Ops+Market_Capitalization.1, data=bankruptcy.df)
summary(cox.out)
#Remove 1s
cox.out <- coxph(Surv(Years_to_Bankrupt,Bankruptcy)~Total_Revenue+Gross_Profit+
Net_Debt.EBITDA+Current_Ratio+Debt_to_Cash_Flow_From_Ops+Market_Capitalization.1, data=bankruptcy.df)
summary(cox.out)
cox.out <- coxph(Surv(Years_to_Bankrupt,Bankruptcy)~Gross_Profit+
Net_Debt.EBITDA+Current_Ratio+Debt_to_Cash_Flow_From_Ops+Market_Capitalization.1, data=bankruptcy.df)
summary(cox.out)
cox.out <- coxph(Surv(Years_to_Bankrupt,Bankruptcy)~Net_Debt.EBITDA+Debt_to_Cash_Flow_From_Ops+Market_Capitalization.1, data=bankruptcy.df)
summary(cox.out)
```