Bankruptcy

Jack Dumanski, Maddie Geittmann, Gabriela Galan

About our Dataset

- Binary response variable: Bankruptcy
- Several ratio and rate variables
- Majority of variables range between 0 and 1
    - Difficult to compare these to non-ratio variables

Taiwanese Bankruptcy Prediction

Donated on 6/27/2020

The data were collected from the Taiwan Economic Journal for the years 1999 to 2009. Company bankruptcy was defined based on the business regulations of the Taiwan Stock Exchange.

Dataset Characteristics	Subject Area	Associated Tasks
Multivariate	Business	Classification
Feature Type	**# Instances**	**# Features**
Integer	6819	95

Dataset Information

Has Missing Values?

No

Our Strategy

Logistic Regression

- → Why?
- → How?

Split into Groups

- → Profitability
- → Liquidity
- → Leverage
- → Growth Rates

Check Accuracy of Models

- → Calculate AIC, ROC, Chi-Square
- → Determine significance of each variable
- → Combine into full model to check significance in larger context

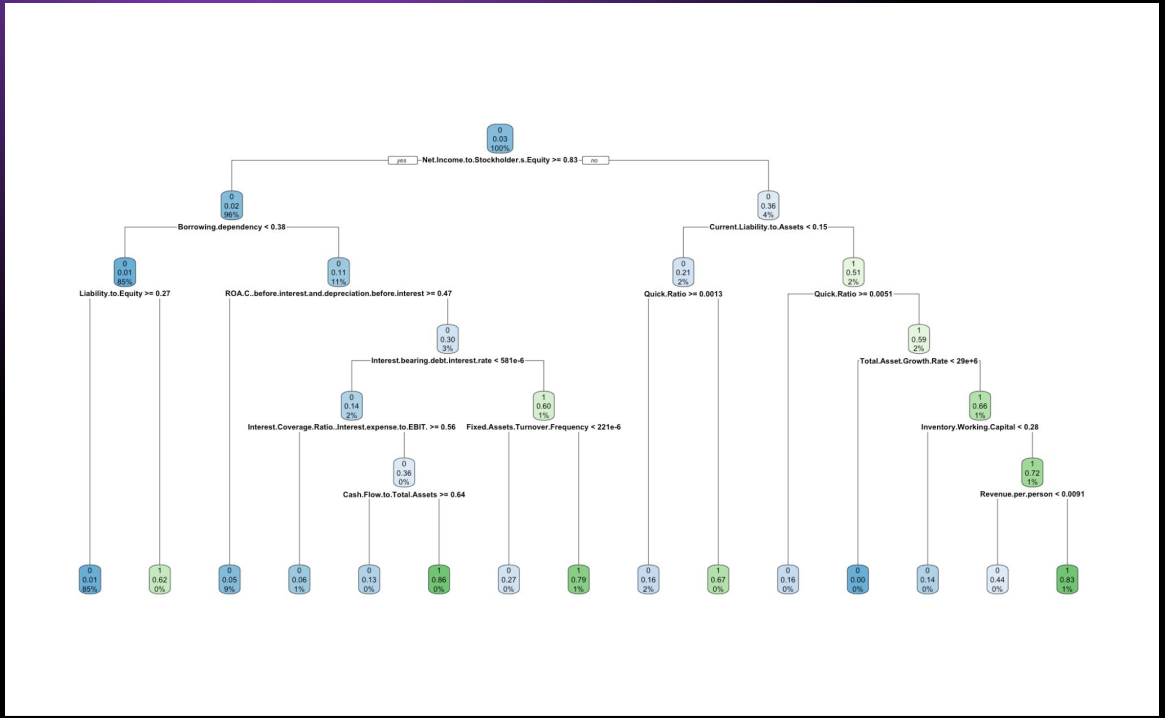
THREATS

- → Collinearity
- → Predictors are inconsistent with ratios

Other potential models for our dataset and drawbacks

Tree model

- Concentrates data into boolean comparisons
- Has 96.3% accuracy in predicting bankruptcy in our data!
- Oversimplification of data, doesn't explain nuances and potential collinearity



Groupings of predictors

01 Profitability

02 Liquidity

03 Leverage

04 Growth Rates

05 Full Model

Profitability

Selected predictors for profitability (11) based on profit or other financial gain:

- Operating gross Margin
- Realized sales gross margin
- Pre-tax ret. interest rate
- **Operating profit growth**
- **Total income/total expense**
- **Cash flow to equity**
- Revenue per Share
- Total asset turnover
- ROA before interest and depreciation before interest rate (3)

Used tidyverse select() function to select these specific columns

Profitability Code: AIC Method

```
40
41  profitability <- bankruptcy %>% #new dataframe with selected columns based on profitability
42    dplyr::select(Bankrupt., Operating.Gross.Margin, Realized.Sales.Gross.Margin, Pre.tax.net.Interest.Rate,
43    Operating.Profit.Growth.Rate, Cash.Flow.to.Equity, Total.income.Total.expense)
44
45  mod_pro
46  summary
47
48  mod_bac
49  summary
    total.i
50
51  mod_for
52  summary                                                                          ), AIC: 1608.2
```

AIC Values	No Step Process	Forward Adding	Backwards Deleting
With Outliers	1608.2	1608.2	1603.1
Without Outliers	1580.7	1580.7	1579.2

```
54  prof_no
55    filter(hatvalues(mod_profit) < 0.03) #removed outliers based on hat values being larger than 0.03
56
57  mod_prof_no_outliers <- glm(Bankrupt. ~., data = prof_no_outliers, family = binomial)
58  summary(mod_prof_no_outliers) #4: intercept, operating.profit.growth.rate, cash.flow.to.equity, total.income.total.expense, AIC: 1580.7
59
60  mod_back_prof <- step(mod_prof_no_outliers, direction = "backward", trace = 0) #step process with backwards deletion
61  summary(mod_back_prof) #4/4 predictors significant (Intercept, operating.profit.growth.rate, cash.flow.to.equity, total.income.total.expense), AIC:
    1579.2
62
63  mod_forward_prof <- step(mod_prof_no_outliers, direction = "forward", trace = 0) #step process with forwards adding
64  summary(mod_forward_prof) #4/7 significant predictors (intercept, operating.profit.growth.rate, cash.flow.to.equity, total.income.total.expense),
    AIC: 1580.7
```

Profitability results

Significant predictors over all 6 models:

Total income/total expenses

Cash flow to equity

Operating profit growth rate

- We will take note of these variables to combine in a larger model later.

Is this model with selected variables significant?

"

At an alpha of 0.05, we reject H0. There is sufficient evidence to suggest that a generalized linear model of Total Income/Expense, Cash Flow/Equity, and Profit Growth Rate without outliers is significant in predicting the log-likelihood of bankruptcy.

```
summary_mod <- glm(Bankrupt. ~ Total.income.Total.expense + Cash.Flow.to.Equity +
Operating.Profit.Growth.Rate, prof_no_outliers, family = binomial)

chi2<-summary(summary_mod)$null-summary(summary_mod)$deviance
p<-pchisq(chi2,df=2,lower.tail=FALSE) #df = num predictors - 1
p

```

```
[1] 1.735546e-78
```

Liquidity

7 explanatory variables

```
94
95  liquidity<- bankruptcy%>%select(Bankrupt.,Current.Ratio,Quick.
    Ratio,Cash.Flow.to.Total.Assets,Cash.Current.Liability,Working
    .Capital.to.Total.Assets,Long.term.fund.suitability.ratio..A.,
    Quick.Assets.Total.Assets)
96
97  mod_liquid<-glm(Bankrupt.~.,data=liquidity,family=binomial)
98  summary(mod_liquid)
99
```

```
Call:
glm(formula = Bankrupt. ~ ., family = binomial, data = liquidity)

Coefficients:
                                          Estimate Std. Error z value Pr(>|z|)
(Intercept)                             1.650e+01  1.513e+00  10.903  < 2e-16 ***
Current.Ratio                          -3.740e-09  1.181e-07  -0.032  0.97473
Quick.Ratio                             2.508e-10  1.426e-10   1.759  0.07859 .
Cash.Flow.to.Total.Assets              -5.025e+00  1.562e+00  -3.217  0.00130 **
Cash.Current.Liability                  2.071e-10  6.327e-11   3.274  0.00106 **
Working.Capital.to.Total.Assets        -2.188e+01  1.650e+00 -13.261  < 2e-16 ***
Long.term.fund.suitability.ratio..A.    2.939e+00  1.277e+00   2.302  0.02133 *
Quick.Assets.Total.Assets               1.352e+00  4.688e-01   2.884  0.00392 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1943.7  on 6818  degrees of freedom
Residual deviance: 1656.9  on 6811  degrees of freedom
AIC: 1672.9

Number of Fisher Scoring iterations: 11
```

Backward and Forward Selection w/ outliers

mod_back<-step(mod_liquid, direction="backward",trace=0)
summary(mod_back)

```
Call:
glm(formula = Bankrupt. ~ Quick.Ratio + Cash.Flow.to.Total.Assets +
    Cash.Current.Liability + Working.Capital.to.Total.Assets +
    Long.term.fund.suitability.ratio..A. + Quick.Assets.Total.Assets,
    family = binomial, data = liquidity)

Coefficients:
                                  Estimate Std. Error z value Pr(>|z|)
(Intercept)                      1.649e+01  1.513e+00  10.902  < 2e-16 ***
Quick.Ratio                      2.510e-10  1.426e-10   1.760  0.07834 .
Cash.Flow.to.Total.Assets       -5.023e+00  1.562e+00  -3.216  0.00130 **
Cash.Current.Liability           2.072e-10  6.327e-11   3.275  0.00105 **
Working.Capital.to.Total.Assets -2.188e+01  1.650e+00 -13.260  < 2e-16 ***
Long.term.fund.suitability.ratio..A.  2.939e+00  1.277e+00   2.302  0.02135 *
Quick.Assets.Total.Assets        1.357e+00  4.685e-01   2.895  0.00379 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1943.7  on 6818  degrees of freedom
Residual deviance: 1657.1  on 6812  degrees of freedom
AIC: 1671.1

Number of Fisher Scoring iterations: 7
```

mod_forward<-step(mod_liquid,direction="forward",trace=0)
summary(mod_forward)

```
Call:
glm(formula = Bankrupt. ~ Current.Ratio + Quick.Ratio + Cash.Flow.to.Total.Assets +
    Cash.Current.Liability + Working.Capital.to.Total.Assets +
    Long.term.fund.suitability.ratio..A. + Quick.Assets.Total.Assets,
    family = binomial, data = liquidity)

Coefficients:
                                  Estimate Std. Error z value Pr(>|z|)
(Intercept)                      1.650e+01  1.513e+00  10.903  < 2e-16 ***
Current.Ratio                   -3.740e-09  1.181e-07  -0.032  0.97473
Quick.Ratio                      2.508e-10  1.426e-10   1.759  0.07859 .
Cash.Flow.to.Total.Assets       -5.025e+00  1.562e+00  -3.217  0.00130 **
Cash.Current.Liability           2.071e-10  6.327e-11   3.274  0.00106 **
Working.Capital.to.Total.Assets -2.188e+01  1.650e+00 -13.261  < 2e-16 ***
Long.term.fund.suitability.ratio..A.  2.939e+00  1.277e+00   2.302  0.02133 *
Quick.Assets.Total.Assets        1.352e+00  4.688e-01   2.884  0.00392 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1943.7  on 6818  degrees of freedom
Residual deviance: 1656.9  on 6811  degrees of freedom
AIC: 1672.9

Number of Fisher Scoring iterations: 11
```

Backward and Forward Selection with a model that has no outliers

```
no_liq_outliers<-liquidity%>%filter(hatvalues(mod_liquid)<0.05)
mod_no_outliers_liq<-glm(Bankrupt.~.,no_liq_outliers,family=binomial)
```

```
Call:
glm(formula = Bankrupt. ~ Cash.Flow.to.Total.Assets + Cash.Current.Liability +
    Working.Capital.to.Total.Assets + Quick.Assets.Total.Assets,
    family = binomial, data = no_liq_outliers)

Coefficients:
                               Estimate Std. Error z value Pr(>|z|)
(Intercept)                   1.875e+01  1.703e+00  11.008  < 2e-16 ***
Cash.Flow.to.Total.Assets    -7.600e+00  1.852e+00  -4.104 4.05e-05 ***
Cash.Current.Liability        3.830e-10  7.971e-11   4.805 1.55e-06 ***
Working.Capital.to.Total.Assets -2.280e+01  1.670e+00 -13.647  < 2e-16 ***
Quick.Assets.Total.Assets     1.680e+00  4.713e-01   3.564 0.000366 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1914.9  on 6794  degrees of freedom
Residual deviance: 1620.4  on 6790  degrees of freedom
AIC: 1630.4

Number of Fisher Scoring iterations: 7
```

```
mod_back_liq<-step(mod_no_outliers_liq,direction="backward",trace=0)
summary(mod_back_liq)
```

```
mod_forward_liq<-step(mod_no_outliers_liq,direction="forward",trace=0)
summary(mod_forward_liq)
```

```
Call:
glm(formula = Bankrupt. ~ Current.Ratio + Quick.Ratio + Cash.Flow.to.Total.Assets +
    Cash.Current.Liability + Working.Capital.to.Total.Assets +
    Long.term.fund.suitability.ratio..A. + Quick.Assets.Total.Assets,
    family = binomial, data = no_liq_outliers)

Coefficients:
                                  Estimate Std. Error z value Pr(>|z|)
(Intercept)                      1.732e+01  2.066e+00   8.385  < 2e-16 ***
Current.Ratio                   -2.657e+01  2.323e+01  -1.144  0.25272
Quick.Ratio                     -3.831e-09  1.158e-07  -0.033  0.97361
Cash.Flow.to.Total.Assets       -7.818e+00  1.880e+00  -4.159 3.20e-05 ***
Cash.Current.Liability           3.857e-10  8.080e-11   4.774 1.80e-06 ***
Working.Capital.to.Total.Assets -2.043e+01  2.539e+00  -8.046 8.55e-16 ***
Long.term.fund.suitability.ratio..A.  1.233e+00  4.335e+00   0.284  0.77609
Quick.Assets.Total.Assets        1.556e+00  4.771e-01   3.261  0.00111 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1914.9  on 6794  degrees of freedom
Residual deviance: 1618.0  on 6787  degrees of freedom
AIC: 1634

Number of Fisher Scoring iterations: 13
```

Is the overall model statistically significant?

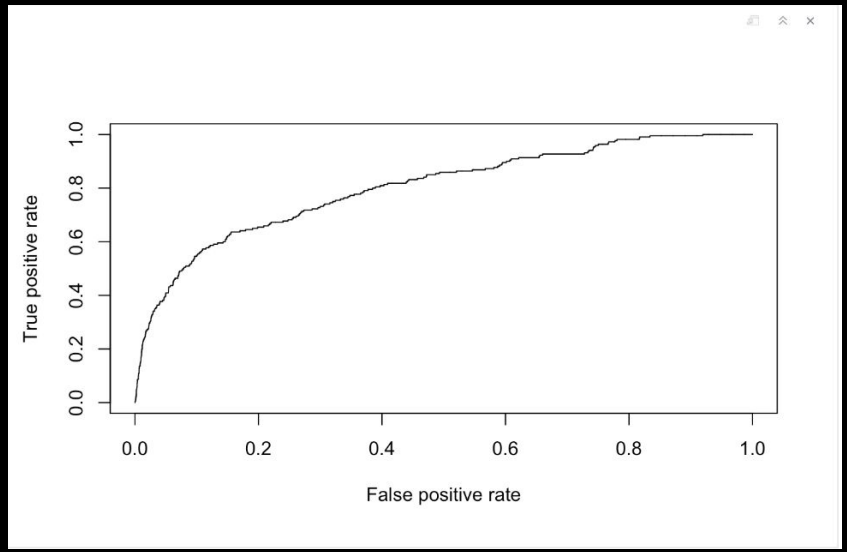
Ho: β1 = · · · = β4 = 0;
Ha: At least one β /= 0

```
mod<-glm(Bankrupt.~ . -Current.Ratio-Quick.Ratio-Long.term.fund.suitability.ratio..A
.,no_liq_outliers,family=binomial)

chi2<-summary(mod)$null-summary(mod)$deviance
chi2
#chi2=2294.4852
p<-pchisq(chi2,df=4,lower.tail=FALSE)
p
#p=1.67619e-62
```

We reject the null at a significant level of 0.05. We can conclude for a logistic regression model predicting the probability of bankruptcy by the intercept, cash.flow.to.total.assets, cash.current.liability,working.capital.to.total.assets,quick.assets.total.assets is statistically significant.

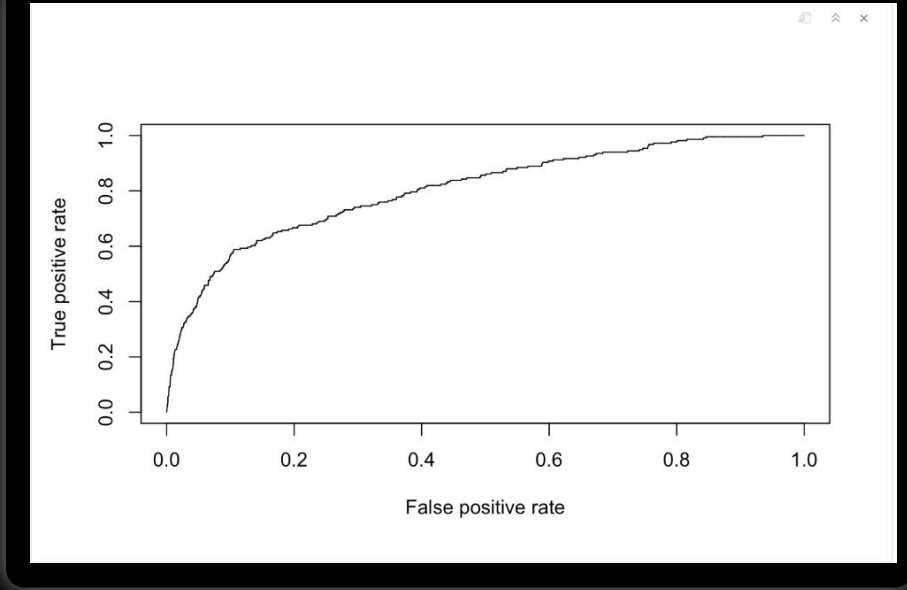
Bonus: ROC Curve

```
library(ROCR)

pred<-predict(mod_back,type="response")
rocpred<-prediction(pred,liquidity$Bankrupt.)
roc<-performance(rocpred,"tpr","fpr")
plot(roc)
```




```
auc<-performance(rocpred,measure="auc")
auc@y.values
mod_back=0.8014513
```

```
pred_mod<-predict(mod,type="response")
rocpred<-prediction(pred_mod,no_liq_outliers$Bankrupt.)
roc<-performance(rocpred,"tpr","fpr")
plot(roc)
auc<-performance(rocpred,measure="auc")
auc@y.values
mod=0.805652
```



Leverage

This model predicts bankruptcy by including variables that measure leverage, or a company's debt to equity ratio.

Explanatory variables include:

- Borrowing dependency
- Contingent liabilities/net worth
- Debt ratio
- Operating funds/Liability
- Interest bearing debt interest rate
- Total debt/total net worth
- Liability/equity
- Equity/liability
- Degree of financial leverage
- Quick assets/current liability

Initial Variable Selection –

Performed backward, forward, and both-direction stepwise variable selection on raw data

Bidirectional & backward stepwise variable selection:
AIC=**1527.6**

```
Call:
glm(formula = Bankrupt. ~ Borrowing.dependency + Contingent.liabilities.Net.worth +
    Debt.ratio.. + Operating.Funds.to.Liability + Interest.bearing.debt.interest.rate +
    Total.debt.Total.net.worth + Liability.to.Equity, family = binomial,
    data = leverage)

Coefficients:
                                  Estimate Std. Error z value Pr(>|z|)
(Intercept)                     -2.559e+00  1.363e+00  -1.877 0.060481 .
Borrowing.dependency             4.178e+01  1.182e+01   3.535 0.000408 ***
Contingent.liabilities.Net.worth 1.518e+01  5.855e+00   2.592 0.009533 **
Debt.ratio..                     2.586e+01  1.661e+00  15.570  < 2e-16 ***
Operating.Funds.to.Liability    -1.676e+01  2.901e+00  -5.777 7.59e-09 ***
Interest.bearing.debt.interest.rate -3.460e-09  1.797e-09  -1.926 0.054164 .
Total.debt.Total.net.worth       1.462e-09  2.918e-10   5.010 5.44e-07 ***
Liability.to.Equity             -5.209e+01  1.573e+01  -3.312 0.000927 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1943.7  on 6818  degrees of freedom
Residual deviance: 1511.6  on 6811  degrees of freedom
AIC: 1527.6

Number of Fisher Scoring iterations: 8
```

Forward stepwise variable selection:

AIC=**1945.7**

```
Call:
glm(formula = Bankrupt. ~ 1, family = binomial, data = leverage)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.40105    0.06853  -49.62   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1943.7  on 6818  degrees of freedom
Residual deviance: 1943.7  on 6818  degrees of freedom
AIC: 1945.7

Number of Fisher Scoring iterations: 6
```

Variable Selection after removing outliers

```
Call:
glm(formula = Bankrupt. ~ Borrowing.dependency + Debt.ratio.. +
    Operating.Funds.to.Liability + Total.debt.Total.net.worth +
    Liability.to.Equity + Liability.Assets.Flag + Equity.to.Liability +
    Quick.Assets.Current.Liability, family = binomial, data = leverage_no_outliers)

Coefficients:
                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                  1.308e-01  3.469e+00   0.038 0.969911
Borrowing.dependency         5.987e+01  1.537e+01   3.895 9.81e-05 ***
Debt.ratio..                 2.772e+01  3.284e+00   8.442  < 2e-16 ***
Operating.Funds.to.Liability -1.778e+01  5.246e+00  -3.389 0.000702 ***
Total.debt.Total.net.worth    2.381e-09  7.808e-10   3.049 0.002293 **
Liability.to.Equity          -8.451e+01  2.386e+01  -3.543 0.000396 ***
Liability.Assets.Flag        -1.410e+01  1.926e+01  -0.732 0.464218
Equity.to.Liability           1.558e+01  4.269e+00   3.648 0.000264 ***
Quick.Assets.Current.Liability -1.090e+02  2.652e+01  -4.111 3.94e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1873.5  on 6790  degrees of freedom
Residual deviance: 1446.0  on 6782  degrees of freedom
AIC: 1464

Number of Fisher Scoring iterations: 9
```

Removing variables with a leverage > 0.05 changed the result of stepwise variable selection.

New AIC=1464

ANOVA & Check for Potential Collinearity

p-value=3.857e-05 when Liability to Equity variable is removed

```
Analysis of Deviance Table

Model 1: Bankrupt. ~ Borrowing.dependency + Debt.ratio.. + Operating.Funds.to.Liability +
    Total.debt.Total.net.worth + Liability.to.Equity + Liability.Assets.Flag +
    Equity.to.Liability + Quick.Assets.Current.Liability
Model 2: Bankrupt. ~ Borrowing.dependency + Debt.ratio.. + Operating.Funds.to.Liability +
    Total.debt.Total.net.worth + Liability.Assets.Flag + Quick.Assets.Current.Liability
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1      6782     1446.0
2      6784     1466.3 -2  -20.326 3.857e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

VIF of remaining explanatory variables all < 5

Borrowing.dependency	Debt.ratio..	Operating.Funds.to.Liability
1.987446	2.280935	1.983316
Total.debt.Total.net.worth	**Liability.Assets.Flag**	**Quick.Assets.Current.Liability**
3.576966	1.059034	2.921925

Check for significance of the revised model in predicting bankruptcy

Ho: β1 = · · · = β6 = 0;
Ha: At least one β /= 0

Chi-squared = 407.1741
P-value = 8.432e-86

We reject the null hypothesis and conclude that the model containing Borrowing dependency, Debt ratio, Operating funds/liability, Total debt/total net worth, and Quick assets/current liability (X1-X5, respectively) is statistically significant in predicting bankruptcy.

$$Bankruptcy = 16.064 * X1 + 17.210 * X2 - 21.267 * X3 + (4.47 * 10^{-}9) * X4 - 126.409 * X5$$

Growth Rates

- We noted several predictors with growth rates and thought it might lead to significant results
- Shows how companies are actively growing/failing to grow

Selected predictors for growth rates (8):

- Realized.Sales.Gross.Profit.Growth.Rate
- Operating.Profit.Growth.Rate
- After.tax.Net.Profit.Growth.Rate
- Regular.Net.Profit.Growth.Rate
- Continuous.Net.Profit.Growth.Rate
- Total.Asset.Growth.Rate
- Net.Value.Growth.Rate
- Total.Asset.Return.Growth.Rate.Ratio

Growth Rates

```
76  growth_rates <- bankruptcy %>%
77    dplyr::select(Bankrupt., contains("growth")) #growth rate category dataframe
78
79  mod_growth <- glm(Bankrupt. ~., data = growth_rates, family = binomial)
80  summary(mod_growth) #3 significant predictors at alpha = 0.05. AIC: 1916.1 (operating.profit.growth.rate,
    total.asset.growth.rate, total.asset.return.growth.rate.ratio)
81
82  growth_back <- step(mod_growth, direction = "backward", trace = 0) #step process with backwards deletion
83  summary(growth_back) #4: operating.profit.growth.rate, after.tax.net.profit.growth.rate,
    total.asset.growth.rate, total.asset.return.growth.rate.ratio, AIC: 1911.2
84  growth_forward <- step(mod_growth, direction = "forward", trace = 0) #step process with forwards adding
85  summary(growth_forward) # 3: operating.profit.growth.rate, total.asset.growth.rate,
    total.asset.return.growth.rate.ratio, AIC: 1916.1
86
87  growth_no_outliers <- growth_rates %>%
88    filter(hatvalues(mod_growth) < 0.05) #removed outliers based on hat values being larger than 0.05
                                                                                                    ds
    Regular.Net.Profit.Growth.Rate, Net.Value.Growth.Rate), 1/9 significant at 0.1
    (Continuous.Net.Profit.Growth.Rate), AIC: 1505.7
98
99
```

AIC Values	No Step	Forwards Adding	Backwards Removing
With Outliers	1916.1	1916.1	1911.2
No Outliers	1505.7	1501.9	1505.7

Profitability results

Significant predictors over all 6 models:

After.tax.Net.Profit.Growth.Rate

Regular.Net.Profit.Growth.Rate,

Net.Value.Growth.Rate

Continuous.Net.Profit.Growth.Rate

Now, with all 4 group's selected variables, we will finally test a full model.

Is this model with these variables significant?

```
98
99  growth_summary <- glm(Bankrupt. ~ After.tax.Net.Profit.Growth.Rate + Regular.Net.Profit.Growth.Rate +
    Net.Value.Growth.Rate + Continuous.Net.Profit.Growth.Rate, growth_no_outliers, family = binomial)
100
101 chi_growth <- summary(growth_summary)$null-summary(growth_summary)$deviance
102 p_growth <- pchisq(chi2,df=3,lower.tail=FALSE) #df = num predictors - 1
103 p_growth
104
105 ```
```

```
[1] 2.627765e-77
```

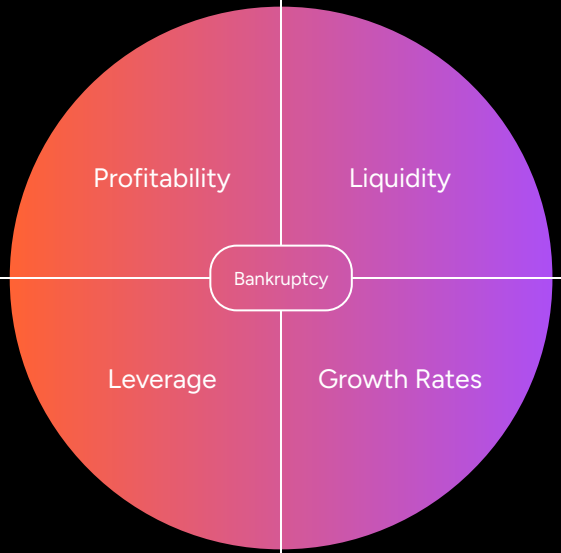
At an alpha of 0.05, we again reject H0. There is sufficient evidence to suggest that a generalized linear model of Total Income/Expense, Cash Flow/Equity, and Profit Growth Rate without outliers is significant in predicting the log-likelihood of bankruptcy.

Predictors from each group

94 → 17 Total Predictors

- → Total income/total expenses
- → Cash flow to equity
- → Operating profit growth rate

- → Intercept
- → Cash.flow.to.total.assets
- → Cash.current.liability
- → Working.capital.to.total.assets
- → Quick.assets.total.assets



- → Debt.ratio.
- → Operating.Funds.to.Liability
- → Total.debt.total.net.worth
- → Equity.to.Liability
- → Quick.Assets.Current.Liability

- → After.tax.net.Profit.Growth.Rate
- → Regular.net.Profit.Growth.Rate
- → Net.Value.Growth.Rate
- → Continuous.Net.Profit.Growth.Rate

Full Model Significance

```
full_model <- glm(Bankrupt. ~ Total.income.Total.expense + Cash.Flow.to.Equity + Operating.Profit.Growth.Rate
+ Cash.Flow.to.Total.Assets + Cash.Current.Liability + Working.Capital.to.Total.Assets +
Quick.Assets.Total.Assets + Debt.ratio.. + Operating.Funds.to.Liability +
                    Liability.Assets.Flag + Equity.to.Liability + Quick.Assets.Current.Liability +
                    After.tax.Net.Profit.Growth.Rate + Regular.Net.Profit.Growth.Rate + Net.Value.Growth.Rate
+ Continuous.Net.Profit.Growth.Rate, bankruptcy, family = binomial)

vif(full_model)

full_model2 <- glm(Bankrupt. ~ Total.income.Total.expense + Cash.Flow.to.Equity +
Operating.Profit.Growth.Rate + Cash.Flow.to.Total.Assets + Cash.Current.Liability +
Working.Capital.to.Total.Assets + Quick.Assets.Total.Assets + Debt.ratio.. + Operating.Funds.to.Liability +
                    Liability.Assets.Flag + Equity.to.Liability + Quick.Assets.Current.Liability +
                    Net.Value.Growth.Rate + Continuous.Net.Profit.Growth.Rate, bankruptcy, family = binomial)
summary(full_model2)

vif(full_model2)
```

- Dropped 2 extreme VIF values from full_model to full_model2, in which every VIF was less than 2, showing little risk of collinearity

AIC: 1283.8

- Original AIC of full model was 16332
- Minimum AIC from grouping predictors was 1464
- Success?

Overall model was significant, p-value of 7.34 * 10^-119 shows we reject H0. At least one predictor was significant in predicting the log-likelihood of bankruptcy.

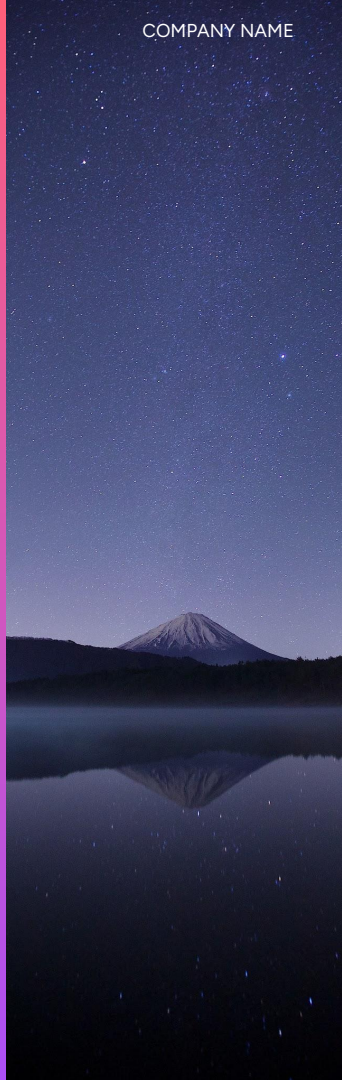
Conclusion

- We were able to minimize AIC through the filtering of predictors with groups, the overall model was significant in predicting bankruptcy in Taiwanese businesses from 1999-2009.

Limitations:

Can’t explain AIC/significance well

Inconsistent variable types



Thank You For Your Time!