ECON-561: Foundation of Analytics

Midterm project

John Garcia

January 9th, 2025

California Lutheran University

Master of Science in Quantitative Economics

Context

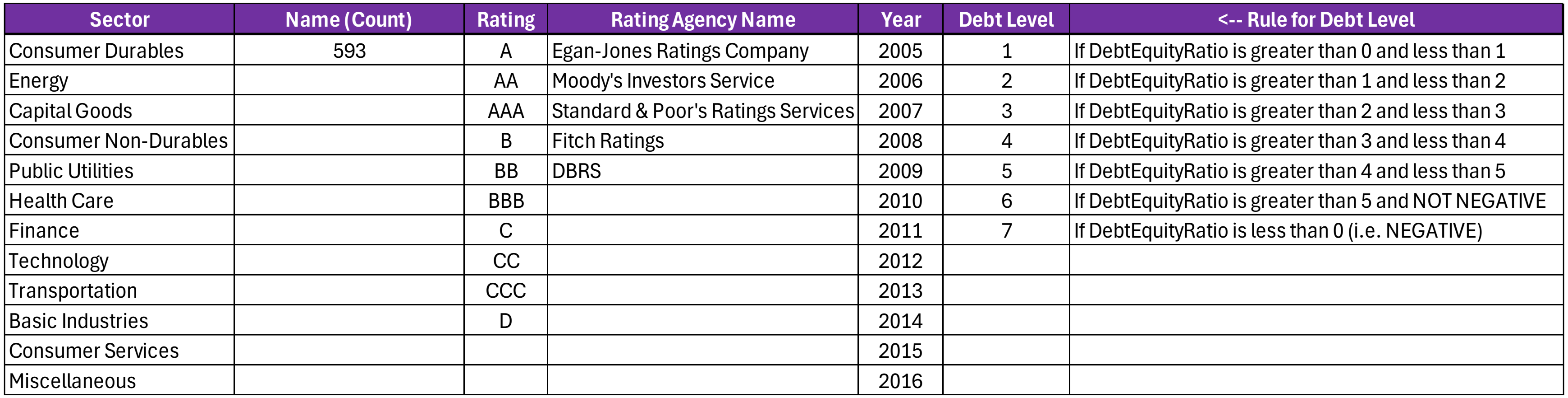
Task:

- Build a model that predicts a firm's 5-year default probability

Resources:

- JMP Pro software
- Historical Credit Rating, Default Probability, and Financial Data on 593 corporations

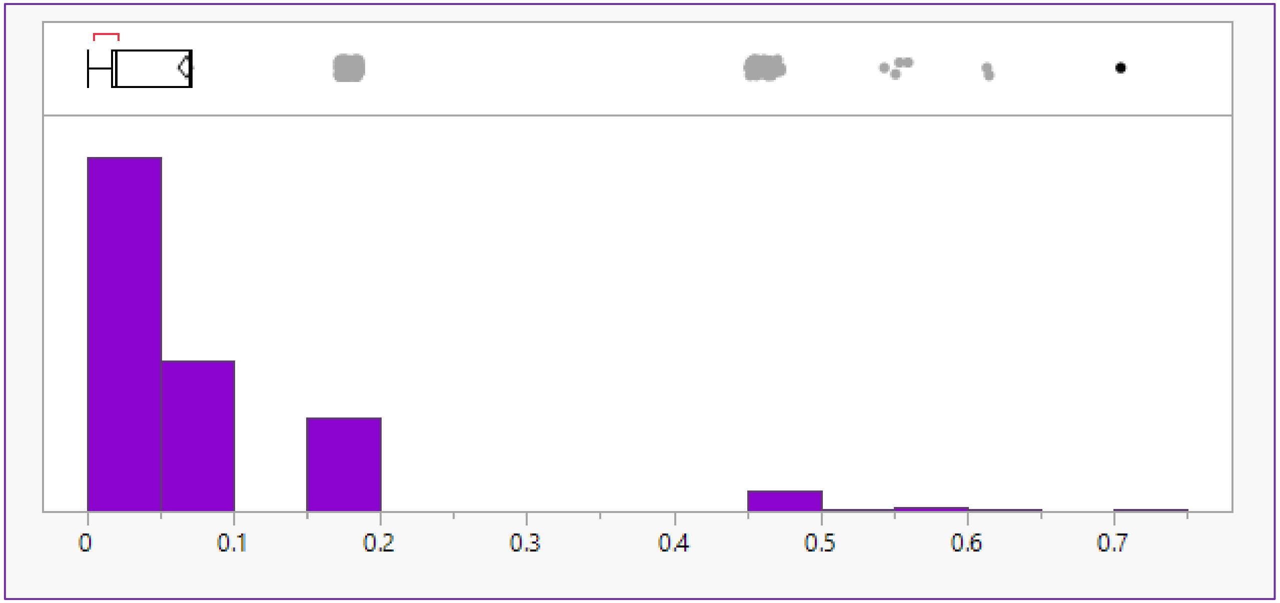
General Details: Categorical Features



Sector	Name (Count)	Rating	Rating Agency Name	Year	Debt Level	<-- Rule for Debt Level
Consumer Durables	593	A	Egan-Jones Ratings Company	2005	1	If DebtEquityRatio is greater than 0 and less than 1
Energy		AA	Moody's Investors Service	2006	2	If DebtEquityRatio is greater than 1 and less than 2
Capital Goods		AAA	Standard & Poor's Ratings Services	2007	3	If DebtEquityRatio is greater than 2 and less than 3
Consumer Non-Durables		B	Fitch Ratings	2008	4	If DebtEquityRatio is greater than 3 and less than 4
Public Utilities		BB	DBRS	2009	5	If DebtEquityRatio is greater than 4 and less than 5
Health Care		BBB		2010	6	If DebtEquityRatio is greater than 5 and NOT NEGATIVE
Finance		C		2011	7	If DebtEquityRatio is less than 0 (i.e. NEGATIVE)
Technology		CC		2012		
Transportation		CCC		2013		
Basic Industries		D		2014		
Consumer Services				2015		
Miscellaneous				2016		

General Details: Response Feature

Default_prob_5yr

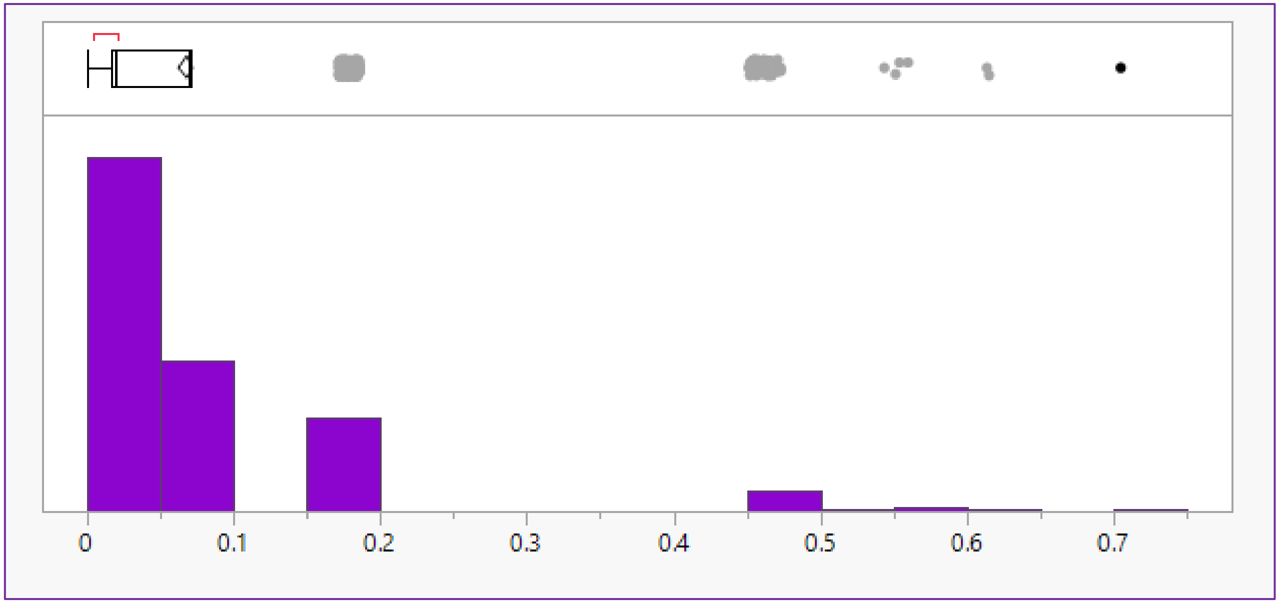


Quantiles										
Minimum				quartile	median	quartile				Maximum
0%	1%	3%	10%	25%	50%	75%	90%	98%	100%	100%
0.00	0.00	0.00	0.00	0.02	0.02	0.07	0.18	0.46	0.47	0.71

Summary Statistics	
Mean	0.067
Median	0.020
Mode	0.020
Minimum	0.001
Maximum	0.705
Range	0.705
Interquartile Range	0.052
5% Trimmed Mean	0.053
Geometric Mean	0.029
Upper 95% Mean	0.072
Lower 95% Mean	0.063
Std Dev	0.098
Std Err Mean	0.002
3*StdDev	0.293
3*StdDev Above Mean	0.360
3*StdDev Below Mean	-0.226
Variance	0.010
Skewness	2.803
Kurtosis	9.071
CV	145.193
N Missing	0
N Zero	0
N Unique	1291.000

General Details: Response Feature

Default_prob_5yr



Quantiles										
Minimum				**quartile**	**median**	**quartile**				**Maximum**
0%	1%	3%	10%	25%	50%	75%	90%	98%	100%	100%
0.00	0.00	0.00	0.00	0.02	0.02	0.07	0.18	0.46	0.47	0.71

Summary Statistics	
Mean	0.067
Median	0.020
Mode	0.020
Minimum	0.001
Maximum	0.705
Range	0.705
Interquartile Range	0.052
5% Trimmed Mean	0.053
Geometric Mean	0.029
Upper 95% Mean	0.072
Lower 95% Mean	0.063
Std Dev	0.098
Std Err Mean	0.002
3*StdDev	0.293
3*StdDev Above Mean	0.360
3*StdDev Below Mean	-0.226
Variance	0.010
Skewness	2.803
Kurtosis	9.071
CV	145.193
N Missing	0
N Zero	0
N Unique	1291.000

General Details: Predictive Features (original)

Cash Flow Features

	freeCashFlowPerShare	cashPerShare	operatingCashFlowPerShare	operatingCashFlowSalesRatio	companyEquityMultiplier	ebitPerRevenue	enterpriseValueMultiple	payablesTurnover
Mean	5,094.72	4,227.55	6,515.12	1.45	3.32	0.44	48.29	38.00
Median	2.13	3.69	4.35	0.13	2.65	0.09	9.27	5.76
Mode	10.16		10.84	0.07		0.07		0.00
Min	-4,912.74	-19.15	-11,950.49	-4.46	-2,555.42	-124.34	-3,749.92	-76.66
Max	5,753,379.81	4,786,803.38	6,439,270.41	688.53	2,562.87	309.69	11,153.61	20,314.88
Range	5,758,292.55	4,786,822.53	6,451,220.90	692.99	5,118.29	434.04	14,903.53	20,391.54
St Dev	146,879.41	122,369.79	177,485.26	19.48	87.51	8.98	528.99	758.74
Variance	21,573,560,588.25	14,974,364,421.48	31,501,017,977.70	379.41	7,657.70	80.68	279,828.89	575,681.15
1Q	0.41	1.56	2.35	0.07	2.05	0.03	6.24	2.20
3Q	4.24	8.10	7.32	0.24	3.66	0.15	12.91	9.49
IQR	3.83	6.53	4.97	0.17	1.61	0.12	6.68	7.29
Skewness	33.65	34.00	30.33	25.43	0.27	22.08	13.94	25.90
Degree of Skew	High right Skew	High right Skew	High right Skew	High right Skew	Approx Nor	High right Skew	High right Skew	High right Skew

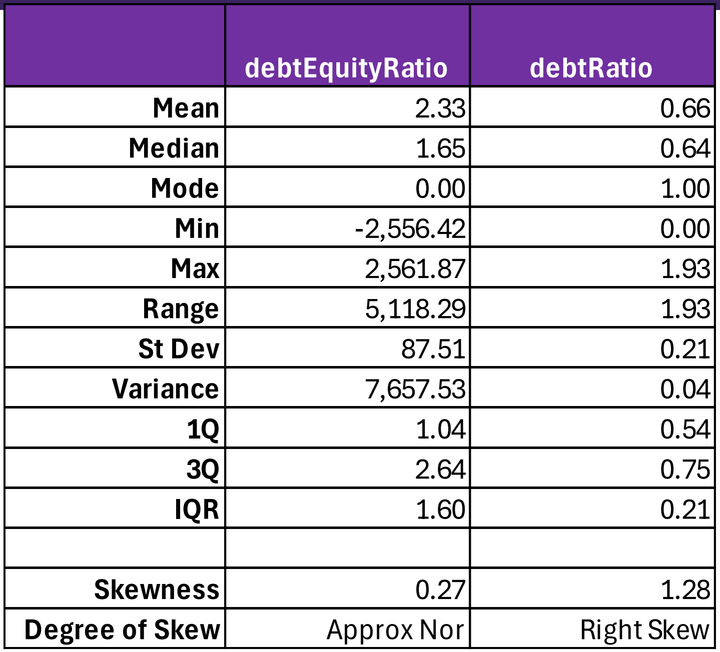
General Details: Predictive Features (original)

Operating Performance features

	currentRatio	returnOnAssets	returnOnEquity	assetTurnover	fixedAssetTurnover
Mean	3.53	-37.52	143.49	3,678.34	7,269.49
Median	1.49	0.05	0.12	0.70	3.81
Mode					
Min	-0.93	-40,213.18	-63.81	-9.16	-26.80
Max	1,725.51	0.49	141,350.21	2,553,148.62	5,156,883.67
Range	1,726.44	40,213.67	141,414.03	2,553,157.77	5,156,910.47
St Dev	44.04	1,165.88	4,405.43	95,630.53	188,950.10
Variance	1,939.65	1,359,287.39	19,407,804.92	9,145,197,644.14	35,702,140,556.05
1Q	1.07	0.02	0.05	0.39	1.02
3Q	2.17	0.08	0.20	1.10	8.52
IQR	1.10	0.06	0.15	0.71	7.50
Skewness	34.31	-32.09	31.68	26.00	26.10
Degree of Skew	High right Skew	High left skew	High right Skew	High right Skew	High right Skew

General Details: Predictive Features (original)

Debt features



	debtEquityRatio	debtRatio
Mean	2.33	0.66
Median	1.65	0.64
Mode	0.00	1.00
Min	-2,556.42	0.00
Max	2,561.87	1.93
Range	5,118.29	1.93
St Dev	87.51	0.21
Variance	7,657.53	0.04
1Q	1.04	0.54
3Q	2.64	0.75
IQR	1.60	0.21
Skewness	0.27	1.28
Degree of Skew	Approx Nor	Right Skew

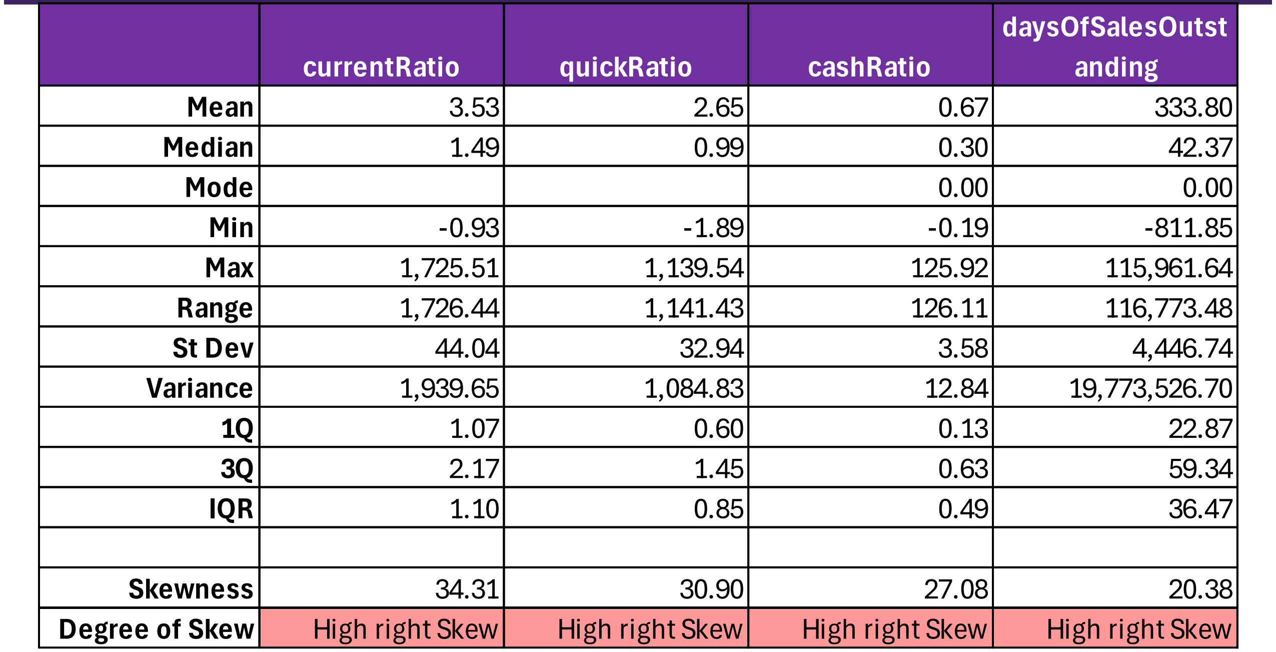
General Details: Predictive Features (original)

Profitability features

	netProfitMargin	pretaxProfitMargin	grossProfitMargin	operatingProfitMargin	returnOnAssets	returnOnCapitalEmployed	returnOnEquity	effectiveTaxRate
Mean	0.28	0.43	0.50	0.59	-37.52	-73.97	143.49	0.40
Median	0.06	0.08	0.41	0.11	0.05	0.07	0.12	0.30
Mode	0.04	0.07	1.00	0.07				0.00
Min	-101.85	-124.34	-14.80	-124.34	-40,213.18	-87,162.16	-63.81	-100.61
Max	198.52	309.69	2.70	410.18	0.49	2.44	141,350.21	429.93
Range	300.36	434.04	17.50	534.53	40,213.67	87,164.60	141,414.03	530.54
St Dev	6.06	8.98	0.53	11.22	1,165.88	2,349.70	4,405.43	10.59
Variance	36.76	80.69	0.28	125.93	1,359,287.39	5,521,073.53	19,407,804.92	112.20
1Q	0.02	0.03	0.23	0.04	0.02	0.03	0.05	0.15
3Q	0.11	0.14	0.85	0.18	0.08	0.14	0.20	0.37
IQR	0.09	0.12	0.62	0.13	0.06	0.11	0.15	0.22
Skewness	17.61	22.08	-14.19	26.47	-32.09	-33.29	31.68	32.28
Degree of Skew	High right Skew	High right Skew	High left skew	High right Skew	High left skew	High left skew	High right Skew	High right Skew

General Details: Predictive Features (original)

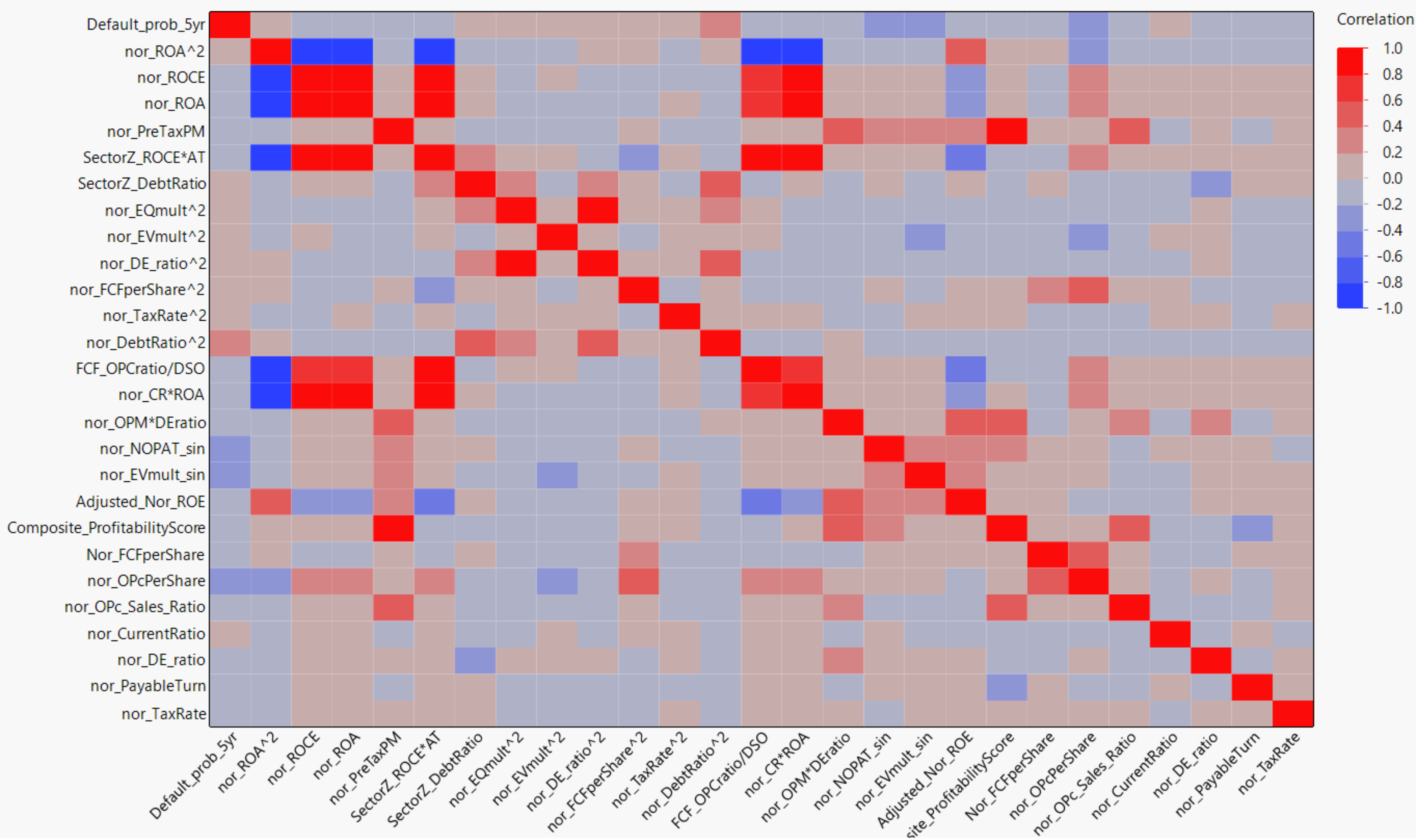
Liquidity features



	currentRatio	quickRatio	cashRatio	daysOfSalesOutstanding
Mean	3.53	2.65	0.67	333.80
Median	1.49	0.99	0.30	42.37
Mode			0.00	0.00
Min	-0.93	-1.89	-0.19	-811.85
Max	1,725.51	1,139.54	125.92	115,961.64
Range	1,726.44	1,141.43	126.11	116,773.48
St Dev	44.04	32.94	3.58	4,446.74
Variance	1,939.65	1,084.83	12.84	19,773,526.70
1Q	1.07	0.60	0.13	22.87
3Q	2.17	1.45	0.63	59.34
IQR	1.10	0.85	0.49	36.47
Skewness	34.31	30.90	27.08	20.38
Degree of Skew	High right Skew	High right Skew	High right Skew	High right Skew

General Details: Predictive Feature

- Correlation Matrix on Selected Predictors and Default Probability does not suggest any OVERLY strong predictor is present in the data.



Data Processing & Feature Engineering

Log-Modular

- Helps Addresses skewed distributions (does not eliminate skewness)
- Preserves sign when dealing with negative values

JMP formula

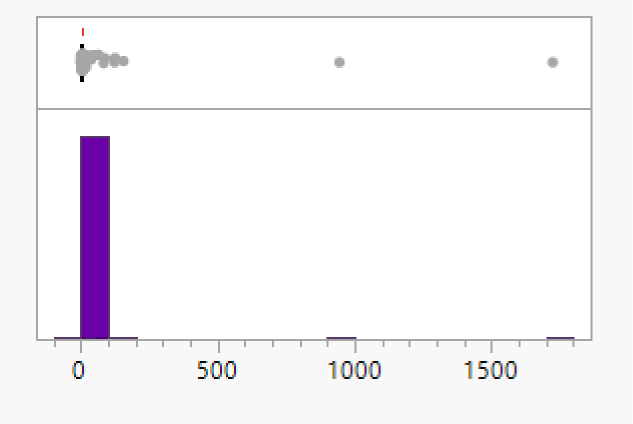
$$if\,(variable > 0\ then\ 1 \cdot \log(|variable| + 1)$$

$$if\,(variable < 0\ then\ -1 \cdot \log(|variable| + 1)$$

Log-Modular transformation applied all 25 original numerical features

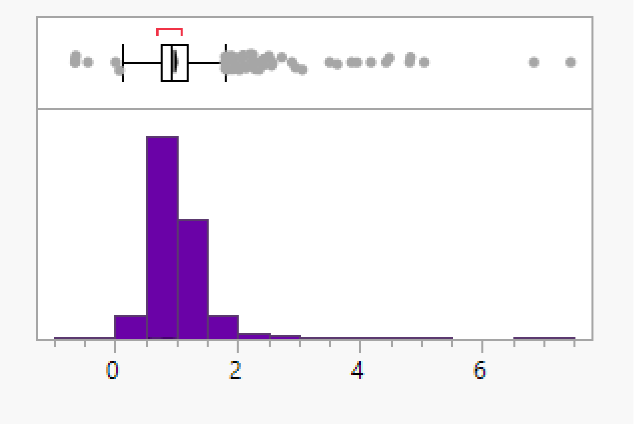
Data Processing & Feature Engineering

Log-Modular example



CurrentRatio

Skewness: 34.33



Nor_CurrentRatio

Skewness: 4.23

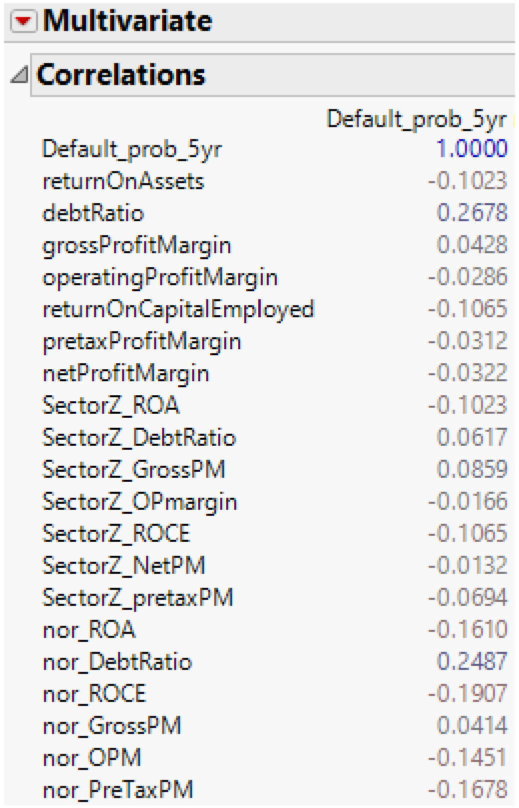
Skewness reduced and distribution is more “Bell Shaped”

Data Processing & Feature Engineering

Did Sector Exhibit "skewness" from Nor Distribution for this feature?

Sector	NetProfit Margin	Pretax Profit margin	gross Profit Margin	Operating Profit Margin	ROA	debtRatio
Basic Industries	Yes	Yes	Yes	Yes	Yes	Yes
Capital Goods	Yes		Yes			y
Consumer Durables			Yes			
Consumer Non-Durables	Yes	Yes		Yes	Yes	
Consumer Services		Yes			Yes	Yes
Energy		Yes		Yes	Yes	Yes
Finance	Yes	Yes	Yes	Yes	Yes	
Health Care	Yes	Yes		Yes	Yes	Yes
Miscellaneous	Yes	Yes		Yes		
Public Utilities	Yes	Yes	Yes	Yes	Yes	Yes
Technology					Yes	Yes
Transportation						Yes

- Certain features received an alternative normalization based on the categorical feature of **"Sector"** attempting to produce potentially better predictive power. (results varied)



Multivariate

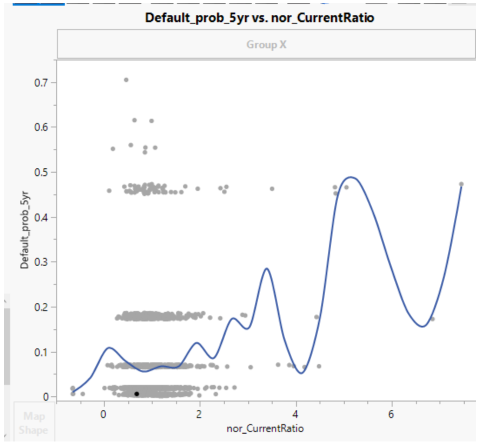
Correlations

	Default_prob_5yr
Default_prob_5yr	1.0000
returnOnAssets	-0.1023
debtRatio	0.2678
grossProfitMargin	0.0428
operatingProfitMargin	-0.0286
returnOnCapitalEmployed	-0.1065
pretaxProfitMargin	-0.0312
netProfitMargin	-0.0322
SectorZ_ROA	-0.1023
SectorZ_DebtRatio	0.0617
SectorZ_GrossPM	0.0859
SectorZ_OPmargin	-0.0166
SectorZ_ROCE	-0.1065
SectorZ_NetPM	-0.0132
SectorZ_pretaxPM	-0.0694
nor_ROA	-0.1610
nor_DebtRatio	0.2487
nor_ROCE	-0.1907
nor_GrossPM	0.0414
nor_OPM	-0.1451
nor_PreTaxPM	-0.1678

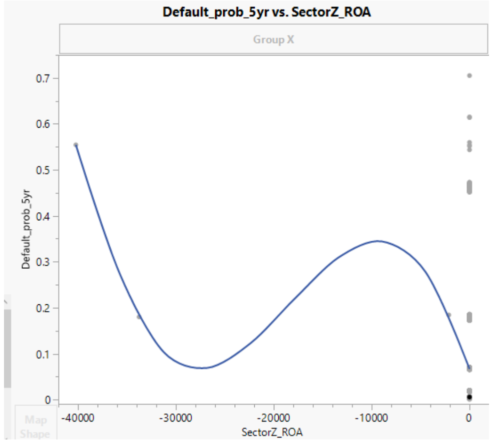
Data Processing & Feature Engineering

- Certain features exhibited non-linear relationships with **Def_prob_5yr** or **over time** were transformed

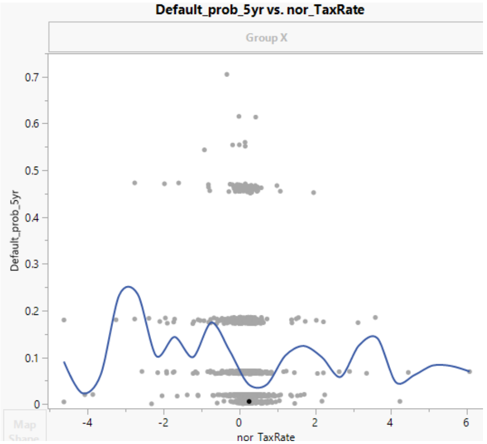
$Nor\_CurrRatio \Rightarrow Nor\_CurrRatio^2$



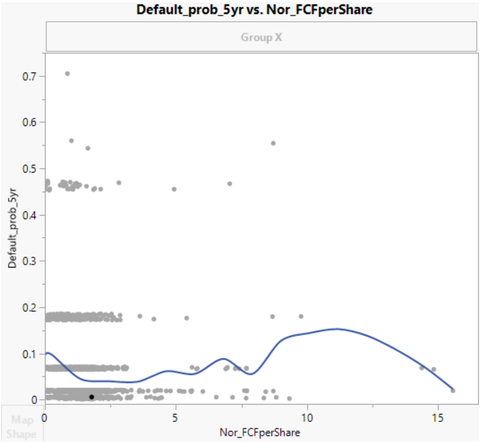
$Nor\_ROA \Rightarrow Nor\_ROA^2$



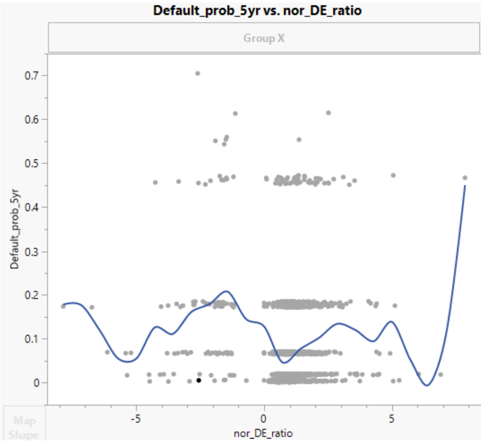
$Nor\_TaxRate \Rightarrow Nor\_TaxRate^2$



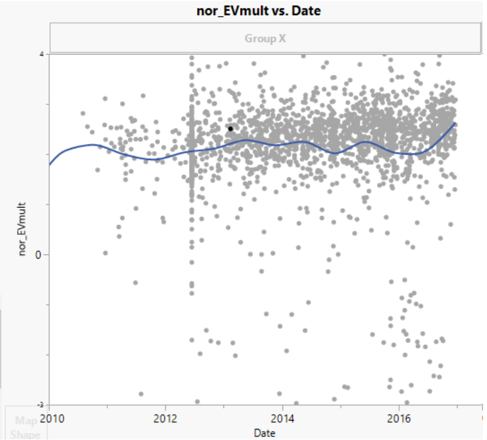
$Nor\_FCFprShr \Rightarrow Nor\_FCFprShr^2$



$Nor\_DE\_ratio \Rightarrow Nor\_DE\_ratio^2$



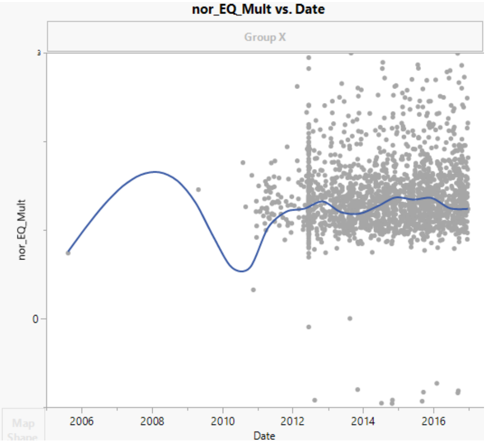
$Nor\ EVmult \Rightarrow \sin(Nor\ EVmult)$



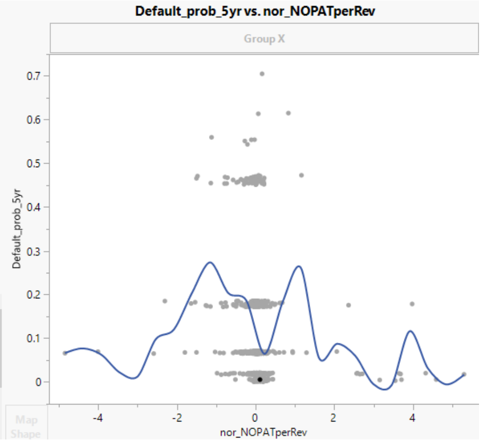
Data Processing & Feature Engineering

- Certain features exhibited non-linear relationships **Def_prob_5yr** or **over time** and were transformed

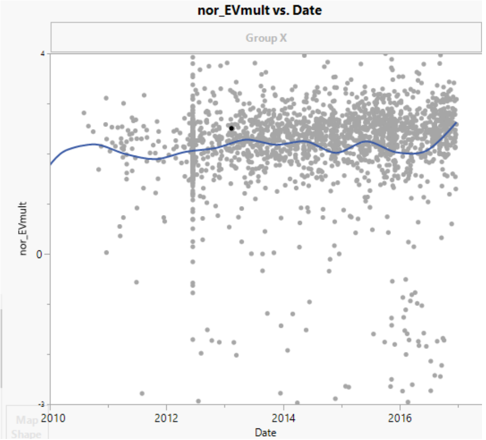
$Nor\ EQmult \Rightarrow \sin(Nor\ EQmult)$



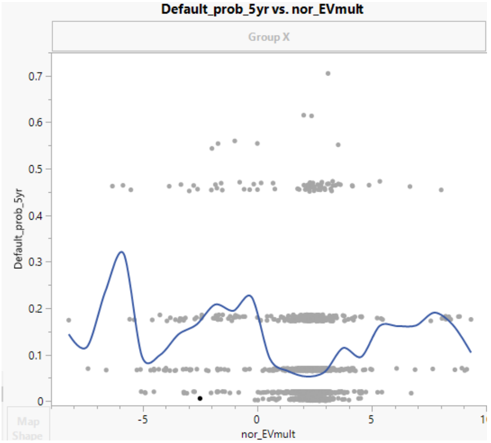
$Nor\ NOPATperRev \Rightarrow \sin(Nor\ NOPATperRev)$



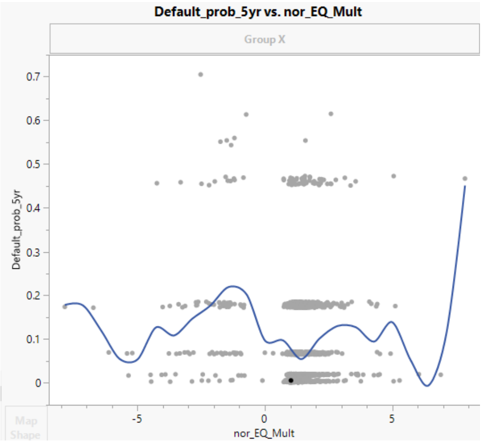
$Nor\ EVmult \Rightarrow Nor\ EVmult^2$



$Nor\ EVmult \Rightarrow Nor\ EVmult^2$



$Nor\ EQmult \Rightarrow Nor\ EQmult^2$



Data Processing & Feature Engineering

- Certain features were interacted, adjusted, or combined

Variables involved		Interaction term created	Intended effect captured
$Nor\ Asset\ turnover$ $\mathrm{SectorZ\_ROCE}$	$\Longrightarrow$	$Nor\ Asset\ turnover * \mathrm{SectorZ\_ROCE}$	A Return on Capital adjusted for Asset utilization
$FCF\ OPcash\ ratio$ $\mathrm{DaysSalesOutstanding}$	$\Longrightarrow$	$\frac{FCF\ OPC\ ratio}{\mathrm{DSO}}$	Cash flow efficiency adjusted by collection speed – and dimension reduction
$nor\ Cash\ Ratio$ $nor\ ROA$	$\Longrightarrow$	$nor\ Cash\ Ratio\ * nor\ ROA$	Interaction of 2 features with Non-linear relationship to Y
$nor\ Op\ Margin$ $nor\ DE\ ratio$	$\Longrightarrow$	$nor\ OPmargin * nor\ DE\ ratio$	Capture the effect of Operating performance relative to leverage
$nor\ Gross\ Margin$ $nor\ Op\ Margin$ $nor\ Net\ Margin$	$\Longrightarrow$	$\frac{(.4 * GrossMargin) + (.35 * OpMargin) + (.25 * NetMargin)}{1}$	Dimension reduction of the mostly none correlation margin variables into 1 variable

Data Processing & Feature Engineering

- Certain features were interacted, adjusted, or combined

Variables involved		Interaction term created	Intended effect captured
Nor Asset turnover SectorZ_ROCE	$\Longrightarrow$	*Nor Asset turnover*$*$ SectorZ_ROCE	A Return on Capital adjusted for Asset utilization
FCF OPcash ratio DaysSalesOutstanding	$\Longrightarrow$	$\frac{FCF\ OPC\ ratio}{\mathrm{DSO}}$	Cash flow efficiency adjusted by collection speed – and dimension reduction
nor Cash Ratio *nor ROA*	$\Longrightarrow$	$nor\ Cash\ Ratio\ * nor\ ROA$	Interaction of 2 features with Non-linear relationship to Y
nor Op Margin *nor DE ratio*	$\Longrightarrow$	$nor\ OPmargin * nor\ DE\ ratio$	Capture the effect of Operating performance relative to leverage
nor Gross Margin *nor Op Margin* *nor Net Margin*	$\Longrightarrow$	$\frac{(.4 * GrossMargin) + (.35 * OpMargin) + (.25 * NetMargin)}{1}$	Dimension reduction of the mostly none correlation margin variables into 1 variable

Direction of correlation for **Cash Ratio** and **OP Margin** is NOT the same across all observations of their interaction term

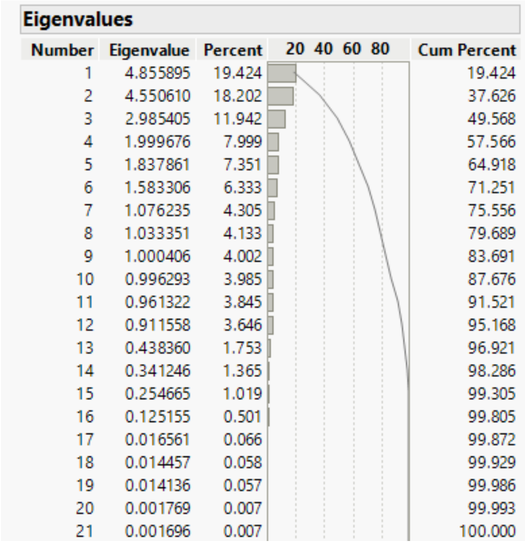
Data Processing & Feature Engineering

- Certain features were interacted, adjusted, or combined

Variables involved		Interaction term created	Intended effect captured
$ebitPerRevenue$ TaxRate	$\Longrightarrow$	$e\text{bitPerRevenue}*(1-\text{TaxRate})$	Stronger profitability measure of Net Operating Profit after Tax (per revenue in this case)
$Nor\ ReturnOnEquity$ $\text{DebtLevel (categorical)}$	$\Longrightarrow$	$\dfrac{Nor\ ROE}{\text{Debt}Level\ scaling\ factor}$	Penalize companies that generate an ROE but have Risky DE ratios (scaling factor increases for companies with DE_ratio >3 or <0

Debt Level	Rule	Scaling Factor Nor_ROE is DIVIDED by:
1	If DebtEquityRatio is greater than 0 and less than 1	1.1
2	If DebtEquityRatio is greater than 1 and less than 2	1.15
3	If DebtEquityRatio is greater than 2 and less than 3	1.2
4	If DebtEquityRatio is greater than 3 and less than 4	1.5
5	If DebtEquityRatio is greater than 4 and less than 5	2
6	If DebtEquityRatio is greater than 5 and NOT NEGATIVE	3
7	If DebtEquityRatio is less than 0 (i.e. NEGATIVE)	4

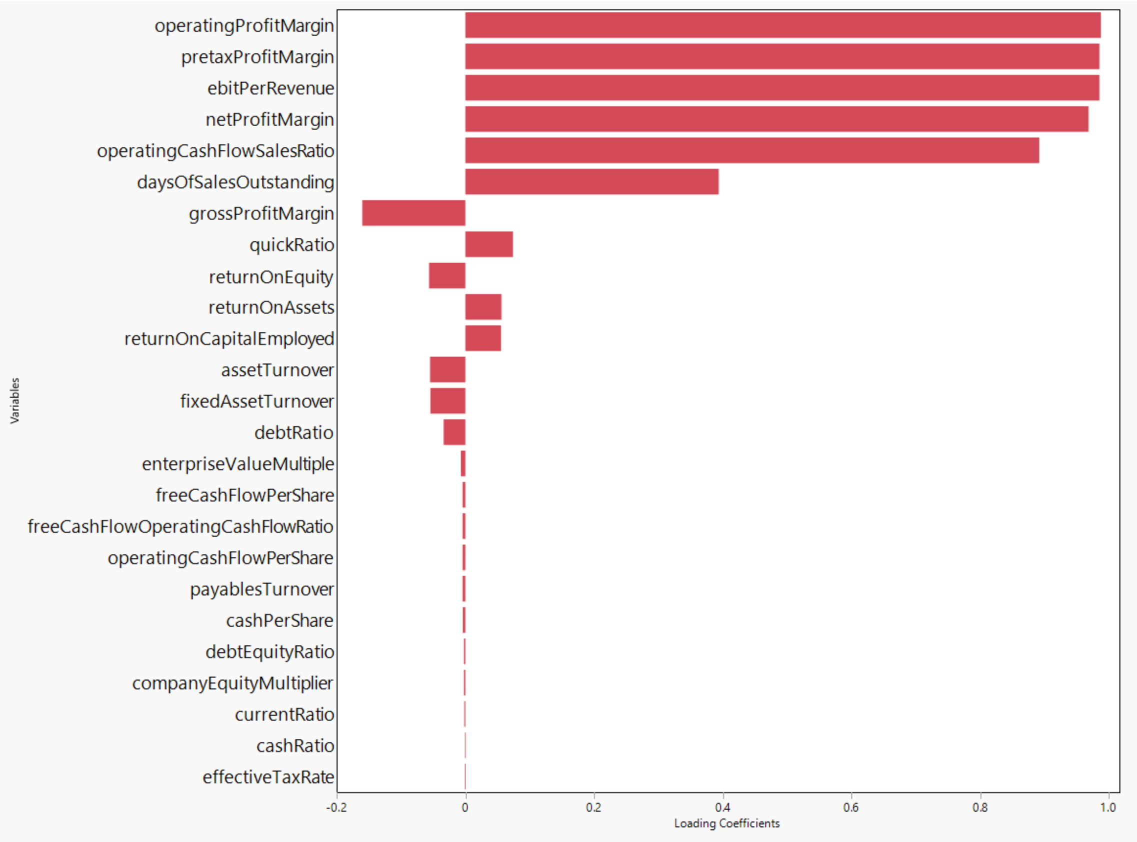
Principle Component Analysis- Original 25 predictors



Eigenvalues

Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	4.855895	19.424		19.424
2	4.550610	18.202		37.626
3	2.985405	11.942		49.568
4	1.999676	7.999		57.566
5	1.837861	7.351		64.918
6	1.583306	6.333		71.251
7	1.076235	4.305		75.556
8	1.033351	4.133		79.689
9	1.000406	4.002		83.691
10	0.996293	3.985		87.676
11	0.961322	3.845		91.521
12	0.911558	3.646		95.168
13	0.438360	1.753		96.921
14	0.341246	1.365		98.286
15	0.254665	1.019		99.305
16	0.125155	0.501		99.805
17	0.016561	0.066		99.872
18	0.014457	0.058		99.929
19	0.014136	0.057		99.986
20	0.001769	0.007		99.993
21	0.001696	0.007		100.000

- 9 latent factors appear in the data (Kaiser Criterion)
- "Margin" variables make significant contribution to PC1 – consider condensing their effect into a smaller dimension
- Variables from EVmultipler downward – consider investigating further and throw-out if necessary



19% of predictor variance is explained by PC1

Most Important variables in PC1

- Op Margin
- PreTax Margin
- ebitPerRevenue
- NetProfitMargin
- OpCash_Sales Ratio

Least Important variables in PC1

- EffectiveTaxRate
- CashRatio
- CurrentRatio
- EquityMultiplier
- DE_ratio

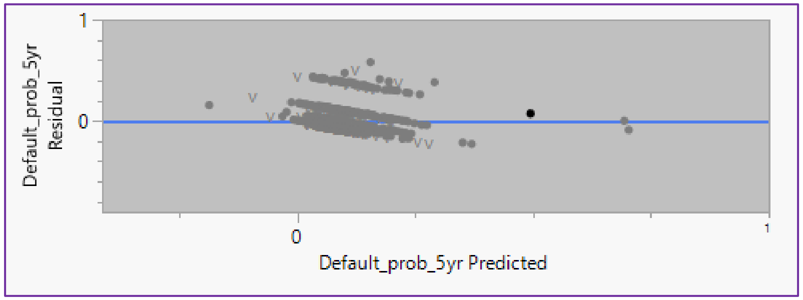
California Lutheran
UNIVERSITY

Model Development - Original 25 predictors

Standard Least Squares

Summary of Fit	
RSquare	0.133
RSquare Adj	0.119
Root Mean Square Error	0.094
Mean of Response	0.069
Observations (or Sum Wgts)	1623

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	25	2.161	0.086	9.762
Error	1597	14.141	0.009	Prob > F
C. Total	1622	16.302		<.0001



Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.047	0.012	-3.98	<.0001
currentRatio	0.000	0.000	0.96	0.338
quickRatio	0.000	0.000	0.21	0.833
cashRatio	0.003	0.001	3.16	0.002
daysOfSalesOutstanding	0.000	0.000	-0.63	0.530
payablesTurnover	0.000	0.000	-1.56	0.120
netProfitMargin	-0.005	0.005	-1.02	0.309
pretaxProfitMargin	-0.035	0.034	-1.02	0.310
grossProfitMargin	0.010	0.005	1.86	0.063
operatingProfitMargin	-0.001	0.003	-0.21	0.831
returnOnAssets	-0.002	0.001	-1.97	0.050
returnOnCapitalEmployed	-0.001	0.001	-1.5	0.134
returnOnEquity	-0.001	0.000	-2.17	0.030
assetTurnover	0.000	0.000	-1.86	0.063
fixedAssetTurnover	0.000	0.000	1.86	0.063
debtEquityRatio	-0.019	0.008	-2.48	0.013
debtRatio	0.132	0.011	11.62	<.0001
effectiveTaxRate	0.000	0.000	-0.19	0.852
freeCashFlowOperatingCashFlowRatio	-0.001	0.001	-1.62	0.104
freeCashFlowPerShare	0.000	0.000	-2.63	0.009
cashPerShare	0.000	0.000	2.64	0.008
companyEquityMultiplier	0.019	0.008	2.49	0.013
ebitPerRevenue	0.039	0.034	1.16	0.247
enterpriseValueMultiple	0.000	0.000	1.23	0.219
operatingCashFlowPerShare	0.000	0.000	2.61	0.009
operatingCashFlowSalesRatio	0.000	0.000	-0.38	0.702

- Model Initial readout has low RMSE, indicating predictions deviate from the actual value 9.4% of the time – not bad right?

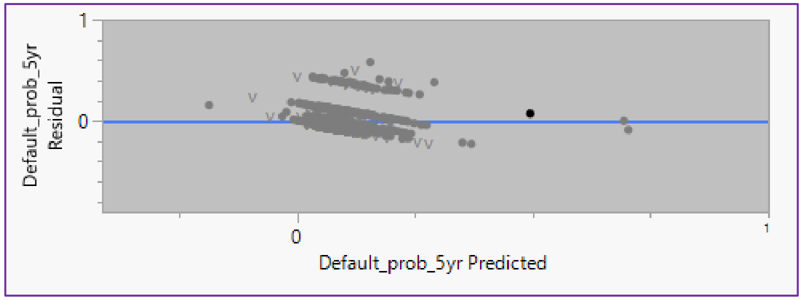
Model Development - Original 25 predictors

Standard Least Squares

Summary of Fit	
RSquare	0.133
RSquare Adj	0.119
Root Mean Square Error	0.094
Mean of Response	0.069
Observations (or Sum Wgts)	1623

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	25	2.161	0.086	9.762
Error	1597	14.141	0.009	Prob > F
C. Total	1622	16.302		<.0001

Source	RSquare	RASE	Freq
Training Set	0.133	0.093	1623
Validation Set	-7.546	0.253	406



Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.047	0.012	-3.98	<.0001
currentRatio	0.000	0.000	0.96	0.338
quickRatio	0.000	0.000	0.21	0.833
cashRatio	0.003	0.001	3.16	0.002
daysOfSalesOutstanding	0.000	0.000	-0.63	0.530
payablesTurnover	0.000	0.000	-1.56	0.120
netProfitMargin	-0.005	0.005	-1.02	0.309
pretaxProfitMargin	-0.035	0.034	-1.02	0.310
grossProfitMargin	0.010	0.005	1.86	0.063
operatingProfitMargin	-0.001	0.003	-0.21	0.831
returnOnAssets	-0.002	0.001	-1.97	0.050
returnOnCapitalEmployed	-0.001	0.001	-1.5	0.134
returnOnEquity	-0.001	0.000	-2.17	0.030
assetTurnover	0.000	0.000	-1.86	0.063
fixedAssetTurnover	0.000	0.000	1.86	0.063
debtEquityRatio	-0.019	0.008	-2.48	0.013
debtRatio	0.132	0.011	11.62	<.0001
effectiveTaxRate	0.000	0.000	-0.19	0.852
freeCashFlowOperatingCashFlowRatio	-0.001	0.001	-1.62	0.104
freeCashFlowPerShare	0.000	0.000	-2.63	0.009
cashPerShare	0.000	0.000	2.64	0.008
companyEquityMultiplier	0.019	0.008	2.49	0.013
ebitPerRevenue	0.039	0.034	1.16	0.247
enterpriseValueMultiple	0.000	0.000	1.23	0.219
operatingCashFlowPerShare	0.000	0.000	2.61	0.009
operatingCashFlowSalesRatio	0.000	0.000	-0.38	0.702

- But the Validation R^2 and RASE are not favorable
- The ***Original*** Linear model does not generalize well.

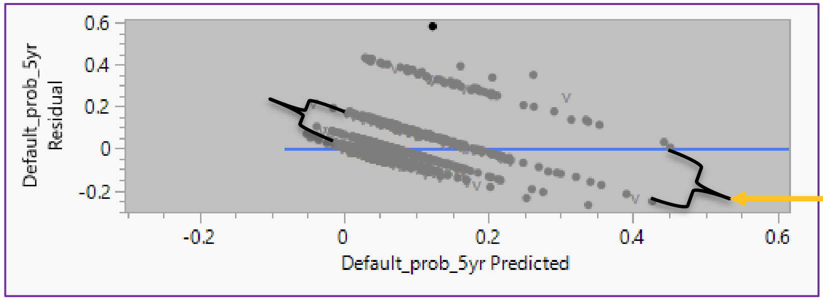
Model Development – Andrew's 26 predictors

Standard Least Squares

Summary of Fit	
RSquare	0.284
RSquare Adj	0.271
Root Mean Square Error	0.086
Mean of Response	0.070
Observations (or Sum Wgts)	1462

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	26	4.221	0.162	21.913
Error	1435	10.631	0.007	Prob > F
C. Total	1461	14.852		<.0001

Source	RSquare	RASE	Freq
Training Set	0.284	0.085	1462
Validation Set	0.204	0.076	367

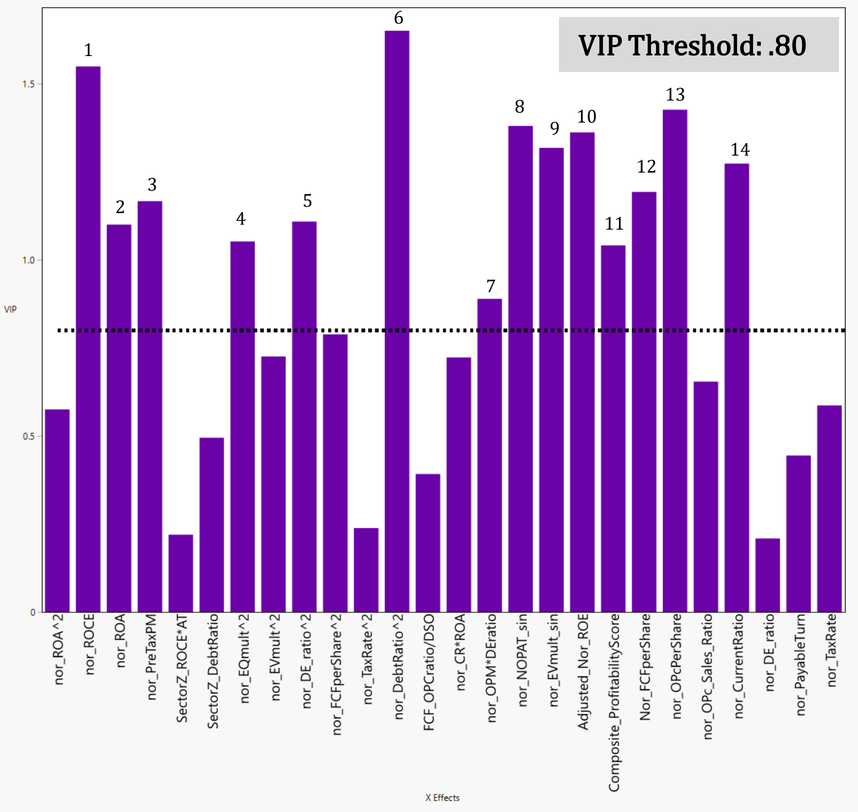


Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.069	0.022	3.11	0.0019
nor_ROA^2	-0.039	0.018	-2.12	0.034
nor_ROCE	0.019	0.038	0.51	0.608
nor_ROA	-0.207	0.079	-2.63	0.009
nor_PreTaxPM	-0.035	0.015	-2.29	0.022
SectorZ_ROCE*AT	0.000	0.009	0.01	0.991
SectorZ_DebtRatio	-0.007	0.005	-1.33	0.182
nor_EQmult^2	-0.022	0.014	-1.59	0.111
nor_EVmult^2	0.001	0.000	2.29	0.022
nor_DE_ratio^2	0.023	0.013	1.71	0.088
nor_FCFperShare^2	0.001	0.000	4.81	<.0001
nor_TaxRate^2	0.001	0.001	0.39	0.694
nor_DebtRatio^2	0.129	0.032	4.06	<.0001
FCF_OPCratio/DSO	-0.002	0.001	-1.77	0.077
nor_CR*ROA	0.040	0.026	1.53	0.127
nor_OPM*DEratio	0.005	0.006	0.86	0.392
nor_NOPAT_sin	-0.158	0.062	-2.56	0.0107
nor_EVmult_sin	-0.018	0.019	-0.96	0.337
Adjusted_Nor_ROE	-0.050	0.025	-2.02	0.044
Composite_ProfitabilityScore	0.052	0.023	2.25	0.024
Nor_FCFperShare	-0.003	0.002	-1.75	0.081
nor_OPcPerShare	-0.013	0.003	-5.35	<.0001
nor_OPc_Sales_Ratio	-0.009	0.008	-1.02	0.306
nor_CurrentRatio	0.032	0.005	6.27	<.0001
nor_DE_ratio	0.001	0.003	0.43	0.670
nor_PayableTurn	-0.002	0.002	-0.97	0.332
nor_TaxRate	-0.008	0.004	-1.91	0.056

- The **Engineered Predictors** enhance the model's predictive ability… but residual vs predicted plot indicates this model is not the right choice.

- Segmentation on plot indicates issues capturing probability values (values bound between 0 and 1)

Model Development – Partial Least Squares

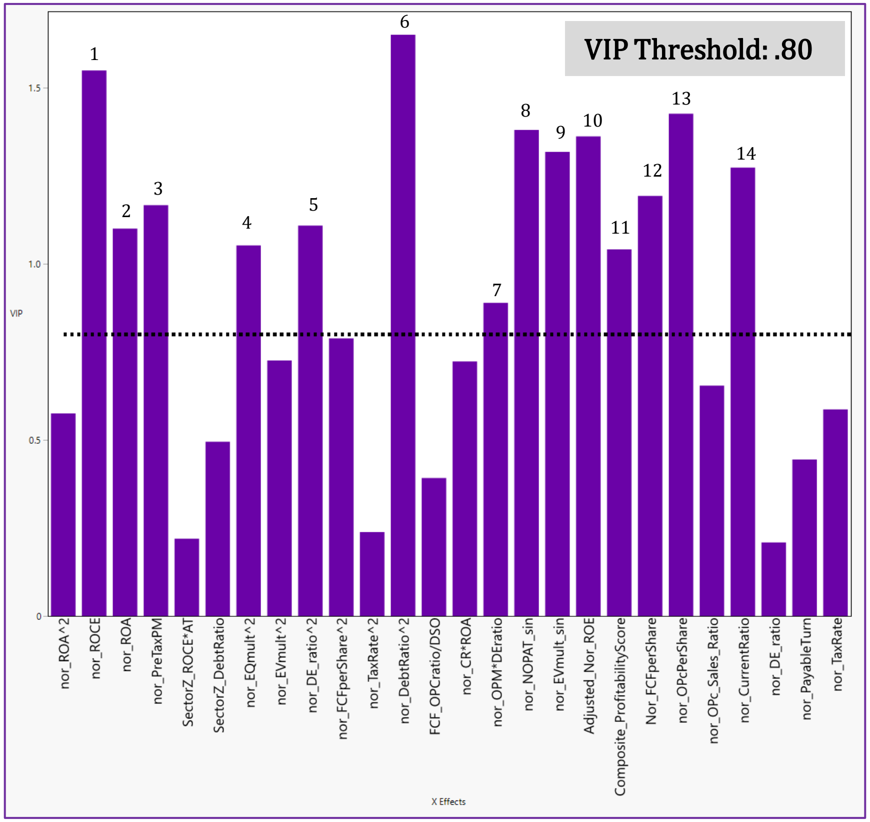


- VIP threshold suggests 14 latent factors are relevant in the data

Number of factors	Root Mean PRESS	R^2X	Cumulative R^2X	R^2Y	Cumulative R^2Y
0	0.846	0.000	0.000	0.000	0.000
1	0.783	0.149	0.149	0.215	0.215
2	0.773	0.084	0.232	0.041	0.257
3	0.756	0.054	0.286	0.016	0.273
4	0.757	0.064	0.350	0.004	0.277
5	0.758	0.055	0.405	0.002	0.279
6	0.755	0.048	0.453	0.001	0.280
7	0.754	0.064	0.517	0.000	0.280
8	0.752	0.043	0.560	0.001	0.281
9	0.752	0.042	0.602	0.001	0.281
10	0.752	0.035	0.637	0.000	0.282
11	0.752	0.032	0.669	0.000	0.282
12	0.753	0.030	0.699	0.000	0.282
13	0.753	0.028	0.727	0.000	0.282
14	0.752	0.024	0.751	0.000	0.283
15	0.750	0.022	0.772	0.000	0.283

- RMPRESS and R^2 values suggest 9 Factors are relevant in the Model.

Model Development – Partial Least Squares



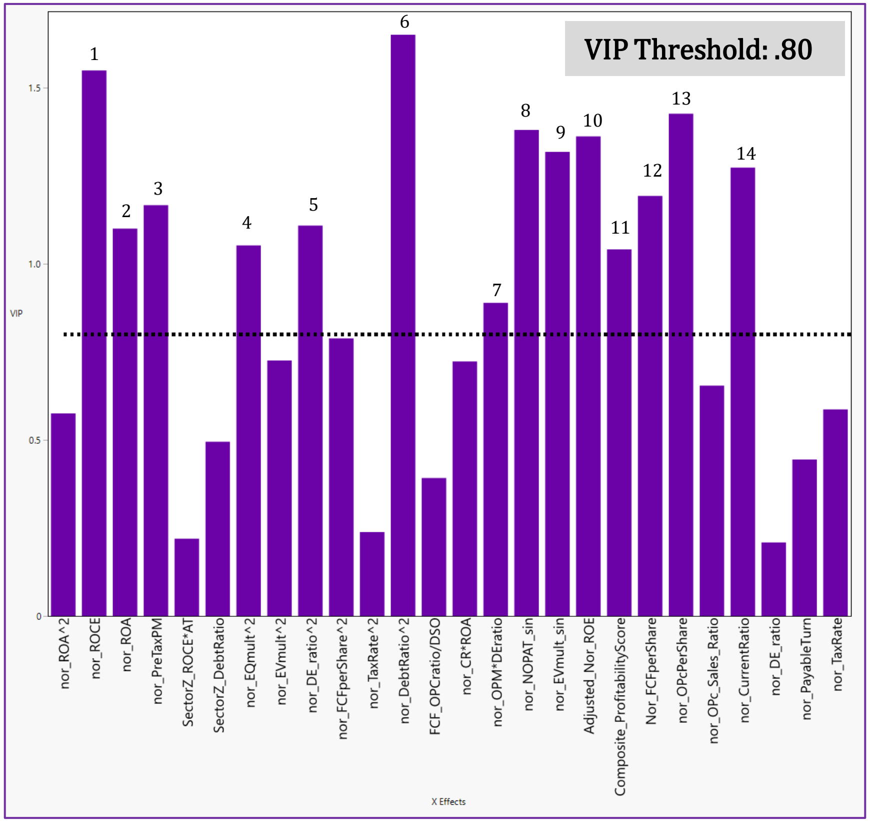
- VIP threshold suggests 14 latent factors are relevant in the data

Number of factors	Root Mean PRESS	R^2X	Cumulative R^2X	R^2Y	Cumulative R^2Y
0	0.846	0.000	0.000	0.000	0.000
1	0.783	0.149	0.149	0.215	0.215
2	0.773	0.084	0.232	0.041	0.257
3	0.756	0.054	0.286	0.016	0.273
4	0.757	0.064	0.350	0.004	0.277
5	0.758	0.055	0.405	0.002	0.279
6	0.755	0.048	0.453	0.001	0.280
7	0.754	0.064	0.517	0.000	0.280
8	0.752	0.043	0.560	0.001	0.281
9	0.752	0.042	0.602	0.001	0.281
10	0.752	0.035	0.637	0.000	0.282
11	0.752	0.032	0.669	0.000	0.282
12	0.753	0.030	0.699	0.000	0.282
13	0.753	0.028	0.727	0.000	0.282
14	0.752	0.024	0.751	0.000	0.283
15	0.750	0.022	0.772	0.000	0.283

- RMPRESS and R^2 values suggest 9 Factors are relevant in the Model.

- Zero variation in Y is explained after adding the 9th factor
- RMPRESS minimized at 15th factor suggestive that this model is fitting more noise in the data than actual data relationships.

Model Development – Partial Least Squares



- VIP threshold suggests 14 latent factors are relevant in the data

Number of factors	Root Mean PRESS	R^2X	Cumulative R^2X	R^2Y	Cumulative R^2Y
0	0.846	0.000	0.000	0.000	0.000
1	0.783	0.149	0.149	0.215	0.215
2	0.773	0.084	0.232	0.041	0.257
3	0.756	0.054	0.286	0.016	0.273
4	0.757	0.064	0.350	0.004	0.277
5	0.758	0.055	0.405	0.002	0.279
6	0.755	0.048	0.453	0.001	0.280
7	0.754	0.064	0.517	0.000	0.280
8	0.752	0.043	0.560	0.001	0.281
9	0.752	0.042	0.602	0.001	0.281
10	0.752	0.035	0.637	0.000	0.282
11	0.752	0.032	0.669	0.000	0.282
12	0.753	0.030	0.699	0.000	0.282
13	0.753	0.028	0.727	0.000	0.282
14	0.752	0.024	0.751	0.000	0.283
15	0.750	0.022	0.772	0.000	0.283

- RMPRESS and R^2 values suggest 9 Factors are relevant in the Model.

- Zero variation in Y is explained after adding the 9th factor
- RMPRESS minimized at 15th factor suggestive that this model is fitting more noise in the data than actual data relationships.

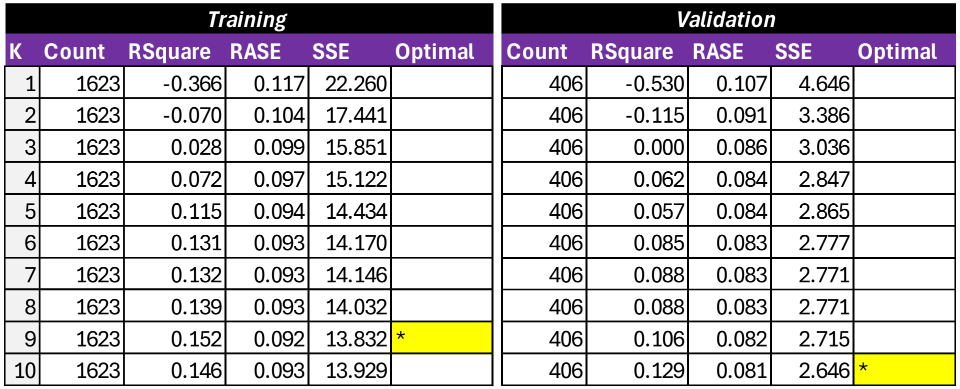
Competitive (not best) RASE

Sum of PLS (Actuals - Predicteds)^2)	Mean of PLS (Actuals - Predicteds)^2)	PLS RASE
12.7495	0.0070	0.0835

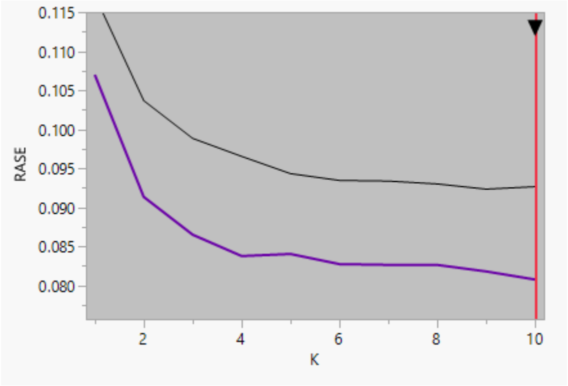
California Lutheran
UNIVERSITY

Model Development – K-nearest neighbors

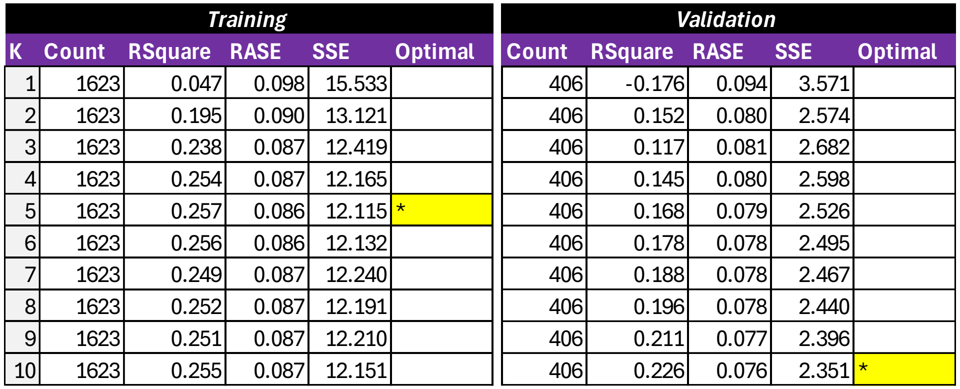
Original 25



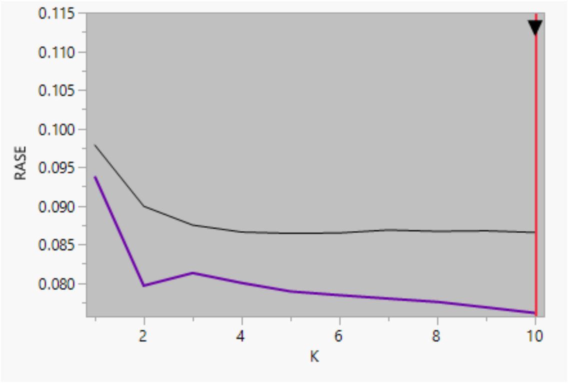
	Training					*Validation*				
K	Count	RSquare	RASE	SSE	Optimal	Count	RSquare	RASE	SSE	Optimal
1	1623	-0.366	0.117	22.260		406	-0.530	0.107	4.646	
2	1623	-0.070	0.104	17.441		406	-0.115	0.091	3.386	
3	1623	0.028	0.099	15.851		406	0.000	0.086	3.036	
4	1623	0.072	0.097	15.122		406	0.062	0.084	2.847	
5	1623	0.115	0.094	14.434		406	0.057	0.084	2.865	
6	1623	0.131	0.093	14.170		406	0.085	0.083	2.777	
7	1623	0.132	0.093	14.146		406	0.088	0.083	2.771	
8	1623	0.139	0.093	14.032		406	0.088	0.083	2.771	
9	1623	0.152	0.092	13.832	*	406	0.106	0.082	2.715	
10	1623	0.146	0.093	13.929		406	0.129	0.081	2.646	*



Andrew's Data



	Training					*Validation*				
K	Count	RSquare	RASE	SSE	Optimal	Count	RSquare	RASE	SSE	Optimal
1	1623	0.047	0.098	15.533		406	-0.176	0.094	3.571	
2	1623	0.195	0.090	13.121		406	0.152	0.080	2.574	
3	1623	0.238	0.087	12.419		406	0.117	0.081	2.682	
4	1623	0.254	0.087	12.165		406	0.145	0.080	2.598	
5	1623	0.257	0.086	12.115	*	406	0.168	0.079	2.526	
6	1623	0.256	0.086	12.132		406	0.178	0.078	2.495	
7	1623	0.249	0.087	12.240		406	0.188	0.078	2.467	
8	1623	0.252	0.087	12.191		406	0.196	0.078	2.440	
9	1623	0.251	0.087	12.210		406	0.211	0.077	2.396	
10	1623	0.255	0.087	12.151		406	0.226	0.076	2.351	*



Takeaways: →

1. Original data set appears to be more stable and better fit under KNN setting
2. Andrew's data appears to be *slightly* better at prediction

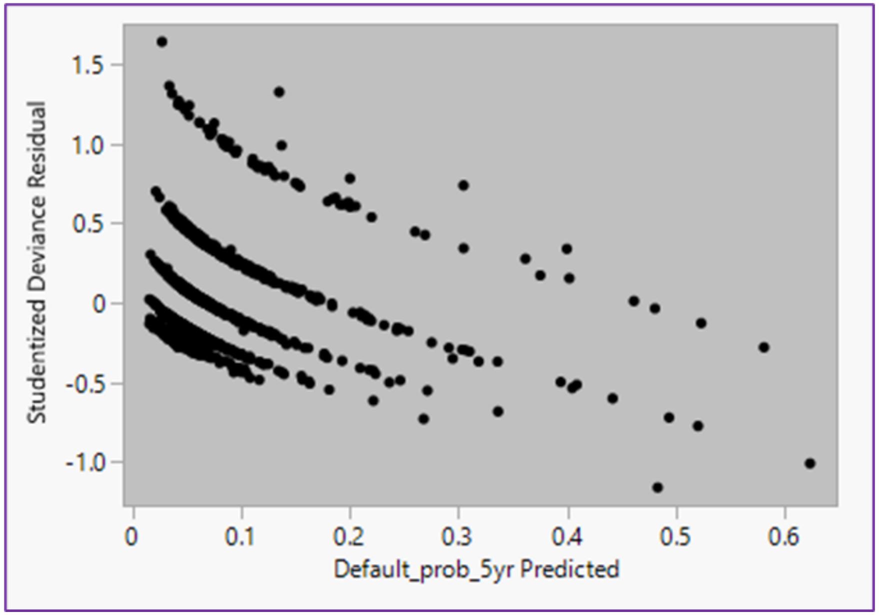
*note: Predicteds for Andrew's Data were saved and tested in generalized regression setting

Model Development – Generalized Linear Model

Source	Logworth		PValue
nor_CurrentRatio	1.073	++	0.08456
nor_OPcPerShare	0.853	++	0.1404
Adjusted_Nor_ROE	0.684	++	0.20689
nor_DebtRatio^2	0.68	+	0.2089
nor_EVmult^2	0.613	+	0.24365
nor_TaxRate	0.44	+	0.36283
nor_OPM*DEratio	0.308	+	0.4925
nor_PreTaxPM	0.27	+	0.53721
SectorZ_DebtRatio	0.255	+	0.5565
nor_FCFperShare^2	0.239	+	0.57687
FCF_OPCratio/DSO	0.23	+	0.58843
nor_ROA^2	0.209		0.61871
nor_ROA	0.193		0.64138
Composite_ProfitabilityScore	0.165		0.68358
nor_OPc_Sales_Ratio	0.111		0.7753
nor_DE_ratio^2	0.108		0.78064
nor_EQmult^2	0.098		0.79723
nor_EVmult_sin	0.097		0.79897
nor_TaxRate^2	0.077		0.83819
nor_ROCE	0.056		0.87884
nor_PayableTurn	0.054		0.88302
nor_DE_ratio	0.045		0.90165
SectorZ_ROCE*AT	0.043		0.90505
nor_NOPAT_sin	0.039		0.91423
nor_CR*ROA	0.037		0.91913
Nor_FCFperShare	0.029		0.93591

	-LogLikelihood	L-R ChiSquare	DF	Prob>Chi Sq
Difference	27.911	55.8212	26	0.0006
Full	329.889			
Reduced	357.800			

Goodness Of Fit Statistic		ChiSquare	DF	Prob>Chi Sq
Pearson		192.0693	1802	1
Deviance		142.5096	1802	1

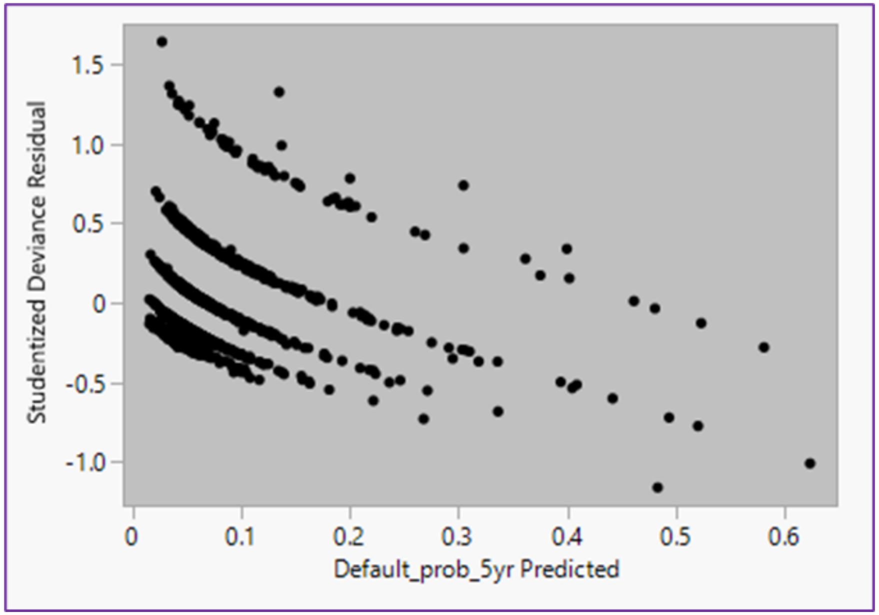


Model Development – Generalized Linear Model

Source	Logworth		PValue
nor_CurrentRatio	1.073	++	0.08456
nor_OPcPerShare	0.853	++	0.1404
Adjusted_Nor_ROE	0.684	++	0.20689
nor_DebtRatio^2	0.68	+	0.2089
nor_EVmult^2	0.613	+	0.24365
nor_TaxRate	0.44	+	0.36283
nor_OPM*DEratio	0.308	+	0.4925
nor_PreTaxPM	0.27	+	0.53721
SectorZ_DebtRatio	0.255	+	0.5565
nor_FCFperShare^2	0.239	+	0.57687
FCF_OPCratio/DSO	0.23	+	0.58843
nor_ROA^2	0.209		0.61871
nor_ROA	0.193		0.64138
Composite_ProfitabilityScore	0.165		0.68358
nor_OPc_Sales_Ratio	0.111		0.7753
nor_DE_ratio^2	0.108		0.78064
nor_EQmult^2	0.098		0.79723
nor_EVmult_sin	0.097		0.79897
nor_TaxRate^2	0.077		0.83819
nor_ROCE	0.056		0.87884
nor_PayableTurn	0.054		0.88302
nor_DE_ratio	0.045		0.90165
SectorZ_ROCE*AT	0.043		0.90505
nor_NOPAT_sin	0.039		0.91423
nor_CR*ROA	0.037		0.91913
Nor_FCFperShare	0.029		0.93591

	-LogLikelihood	L-R ChiSquare	DF	Prob>Chi Sq
Difference	27.911	55.8212	26	0.0006
Full	329.889			
Reduced	357.800			

Goodness Of Fit Statistic		ChiSquare	DF	Prob>Chi Sq
Pearson		192.0693	1802	1
Deviance		142.5096	1802	1



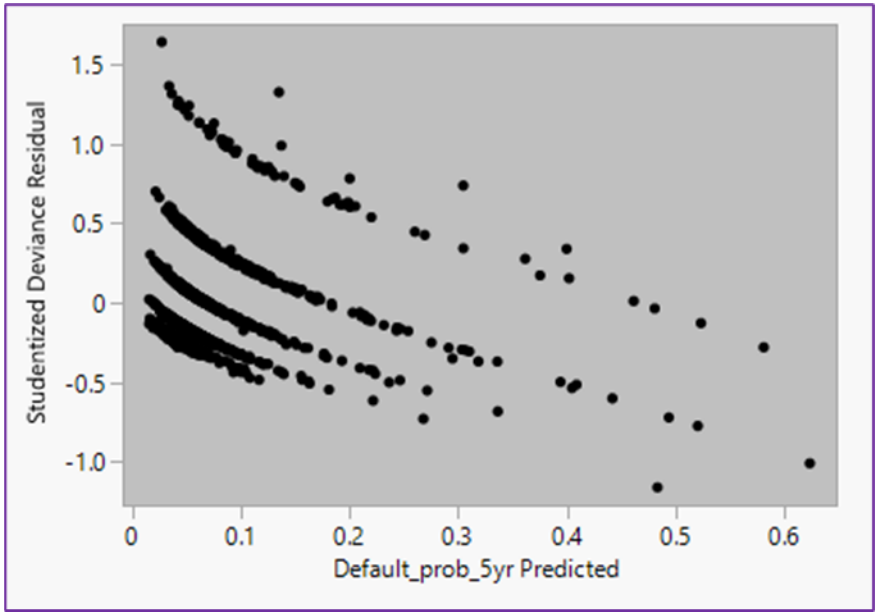
- Only 1 predictor was found to be statistically significant

Model Development – Generalized Linear Model Nor Distribution

Source	Logworth		PValue
nor_CurrentRatio	1.073	++	0.08456
nor_OPcPerShare	0.853	++	0.1404
Adjusted_Nor_ROE	0.684	++	0.20689
nor_DebtRatio^2	0.68	+	0.2089
nor_EVmult^2	0.613	+	0.24365
nor_TaxRate	0.44	+	0.36283
nor_OPM*DEratio	0.308	+	0.4925
nor_PreTaxPM	0.27	+	0.53721
SectorZ_DebtRatio	0.255	+	0.5565
nor_FCFperShare^2	0.239	+	0.57687
FCF_OPCratio/DSO	0.23	+	0.58843
nor_ROA^2	0.209		0.61871
nor_ROA	0.193		0.64138
Composite_ProfitabilityScore	0.165		0.68358
nor_OPc_Sales_Ratio	0.111		0.7753
nor_DE_ratio^2	0.108		0.78064
nor_EQmult^2	0.098		0.79723
nor_EVmult_sin	0.097		0.79897
nor_TaxRate^2	0.077		0.83819
nor_ROCE	0.056		0.87884
nor_PayableTurn	0.054		0.88302
nor_DE_ratio	0.045		0.90165
SectorZ_ROCE*AT	0.043		0.90505
nor_NOPAT_sin	0.039		0.91423
nor_CR*ROA	0.037		0.91913
Nor_FCFperShare	0.029		0.93591

	-LogLikelihood	L-R ChiSquare	DF	Prob>ChiSq
Difference	27.911	55.8212	26	0.0006
Full	329.889			
Reduced	357.800			

Goodness Of Fit Statistic		ChiSquare	DF	Prob>ChiSq
Pearson		192.0693	1802	1
Deviance		142.5096	1802	1



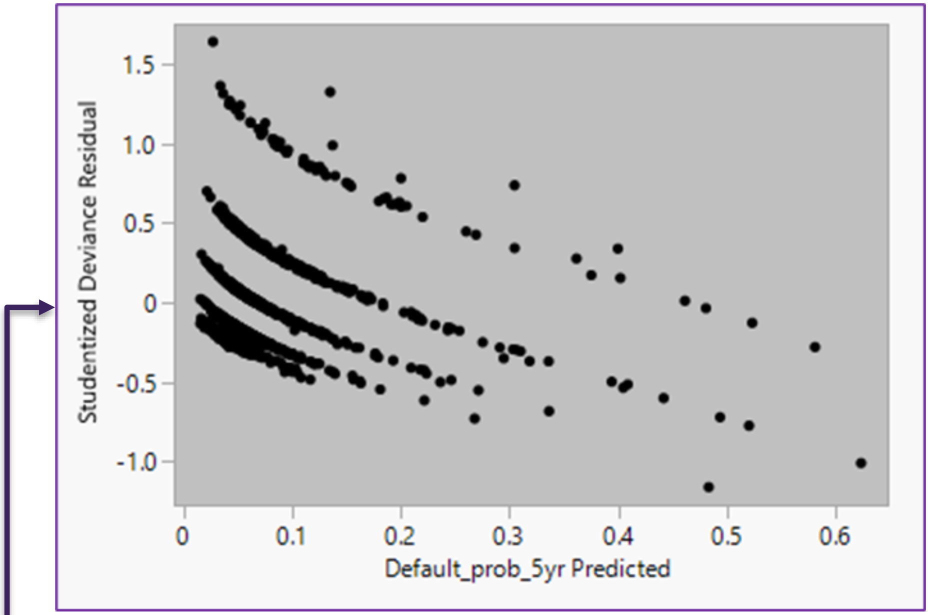
- Difference in Loglikelihood between full model and an “only the intercept” model suggests weak contribution from individual predictors

Model Development – Generalized Linear Model Nor Distribution

Source	Logworth		PValue
nor_CurrentRatio	1.073	++	0.08456
nor_OPcPerShare	0.853	++	0.1404
Adjusted_Nor_ROE	0.684	++	0.20689
nor_DebtRatio^2	0.68	+	0.2089
nor_EVmult^2	0.613	+	0.24365
nor_TaxRate	0.44	+	0.36283
nor_OPM*DEratio	0.308	+	0.4925
nor_PreTaxPM	0.27	+	0.53721
SectorZ_DebtRatio	0.255	+	0.5565
nor_FCFperShare^2	0.239	+	0.57687
FCF_OPCratio/DSO	0.23	+	0.58843
nor_ROA^2	0.209		0.61871
nor_ROA	0.193		0.64138
Composite_ProfitabilityScore	0.165		0.68358
nor_OPc_Sales_Ratio	0.111		0.7753
nor_DE_ratio^2	0.108		0.78064
nor_EQmult^2	0.098		0.79723
nor_EVmult_sin	0.097		0.79897
nor_TaxRate^2	0.077		0.83819
nor_ROCE	0.056		0.87884
nor_PayableTurn	0.054		0.88302
nor_DE_ratio	0.045		0.90165
SectorZ_ROCE*AT	0.043		0.90505
nor_NOPAT_sin	0.039		0.91423
nor_CR*ROA	0.037		0.91913
Nor_FCFperShare	0.029		0.93591

	-LogLikelihood	L-R ChiSquare	DF	Prob>Chi Sq
Difference	27.911	55.8212	26	0.0006
Full	329.889			
Reduced	357.800			

Goodness Of Fit Statistic		ChiSquare	DF	Prob>Chi Sq
Pearson		192.0693	1802	1
Deviance		142.5096	1802	1



- Difference in Loglikelihood between full model and an "only the intercept" model suggests weak contribution from individual predictors

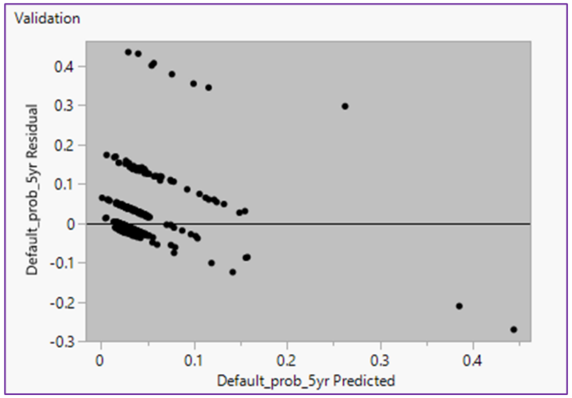
California Lutheran UNIVERSITY

Model Development – Generalized Regression Beta distribution

Response Distribution	Estimation Method	Validation Method	Nonzero Parameters	AICc	BIC	Generalized RSquare	Generalized RSquare
Beta	Maximum Likelihood	Validation Column	28	1889.4779	2036.396	-101.327	-34.040

Term	Estimate	Std Error	Wald ChiSquare	Prob > ChiSquare
Intercept	-3.863	0.127	932.562	<.0001
nor_ROA^2	0.367	0.287	1.634	0.201
nor_ROCE	1.373	0.139	98.225	<.0001
nor_ROA	-3.285	0.389	71.423	<.0001
nor_PreTaxPM	-0.238	0.115	4.261	0.039
SectorZ_ROCE*AT	-0.299	0.049	37.430	<.0001
SectorZ_DebtRatio	-0.140	0.034	17.253	<.0001
nor_EQmult^2	-0.057	0.055	1.059	0.303
nor_EVmult^2	0.016	0.001	200.673	<.0001
nor_DE_ratio^2	0.080	0.055	2.148	0.143
nor_FCFperShare^2	-0.001	0.003	0.183	0.669
nor_TaxRate^2	0.027	0.008	12.339	0.000
nor_DebtRatio^2	1.856	0.143	167.596	<.0001
FCF_OPCratio/DSO	-0.018	0.012	2.221	0.136
nor_CR*ROA	1.065	0.000	.	.
nor_OPM*DEratio	0.158	0.034	21.212	<.0001
nor_NOPAT_sin	-0.679	0.395	2.961	0.085
nor_EVmult_sin	-0.229	0.095	5.773	0.016
Adjusted_Nor_ROE	-1.015	0.145	49.035	<.0001
Composite_ProfitabilityScore	0.090	0.156	0.335	0.563
Nor_FCFperShare	-0.068	0.015	20.735	<.0001
nor_OPcPerShare	-0.136	0.017	67.462	<.0001
nor_OPc_Sales_Ratio	-0.078	0.064	1.499	0.221
nor_CurrentRatio	0.433	0.020	446.290	<.0001
nor_DE_ratio	-0.048	0.011	18.667	<.0001
nor_PayableTurn	0.001	0.017	0.001	0.975
nor_TaxRate	-0.088	0.026	11.348	0.001

Measure	Training	Validation
Number of rows	1462	367
Sum of Frequencies	1462	367
-LogLikelihood	916.172	-14.055
Number of Parameters	28	28
BIC	2036.396	137.240
AICc	1889.478	32.694
Generalized RSquare	-101.327	-34.040

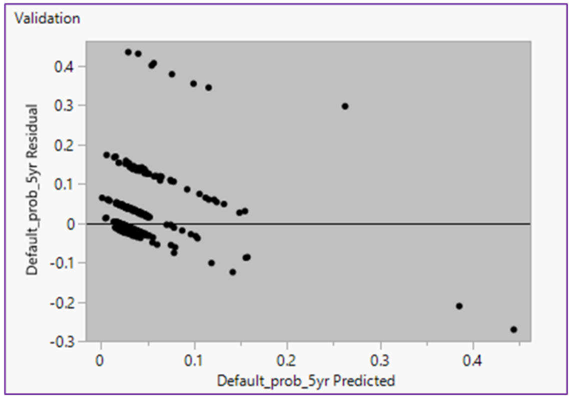


Model Development – Generalized Regression Beta distribution

Response Distribution	Estimation Method	Validation Method	Nonzero Parameters	AICc	BIC	Generalized RSquare	Generalized RSquare
Beta	Maximum Likelihood	Validation Column	28	1889.4779	2036.396	-101.327	-34.040

Term	Estimate	Std Error	Wald ChiSquare	Prob > ChiSquare
Intercept	-3.863	0.127	932.562	<.0001
nor_ROA^2	0.367	0.287	1.634	0.201
nor_ROCE	1.373	0.139	98.225	<.0001
nor_ROA	-3.285	0.389	71.423	<.0001
nor_PreTaxPM	-0.238	0.115	4.261	0.039
SectorZ_ROCE*AT	-0.299	0.049	37.430	<.0001
SectorZ_DebtRatio	-0.140	0.034	17.253	<.0001
nor_EQmult^2	-0.057	0.055	1.059	0.303
nor_EVmult^2	0.016	0.001	200.673	<.0001
nor_DE_ratio^2	0.080	0.055	2.148	0.143
nor_FCFperShare^2	-0.001	0.003	0.183	0.669
nor_TaxRate^2	0.027	0.008	12.339	0.000
nor_DebtRatio^2	1.856	0.143	167.596	<.0001
FCF_OPCratio/DSO	-0.018	0.012	2.221	0.136
nor_CR*ROA	1.065	0.000	.	.
nor_OPM*DEratio	0.158	0.034	21.212	<.0001
nor_NOPAT_sin	-0.679	0.395	2.961	0.085
nor_EVmult_sin	-0.229	0.095	5.773	0.016
Adjusted_Nor_ROE	-1.015	0.145	49.035	<.0001
Composite_ProfitabilityScore	0.090	0.156	0.335	0.563
Nor_FCFperShare	-0.068	0.015	20.735	<.0001
nor_OPcPerShare	-0.136	0.017	67.462	<.0001
nor_OPc_Sales_Ratio	-0.078	0.064	1.499	0.221
nor_CurrentRatio	0.433	0.020	446.290	<.0001
nor_DE_ratio	-0.048	0.011	18.667	<.0001
nor_PayableTurn	0.001	0.017	0.001	0.975
nor_TaxRate	-0.088	0.026	11.348	0.001

Measure	Training	Validation
Number of rows	1462	367
Sum of Frequencies	1462	367
-LogLikelihood	916.172	-14.055
Number of Parameters	28	28
BIC	2036.396	137.240
AICc	1889.478	32.694
Generalized RSquare	-101.327	-34.040



- Alternate distribution of Response variable uncovers the individual predictors contribution

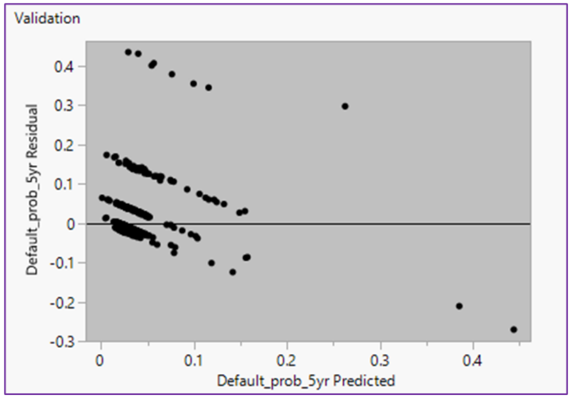
California Lutheran
UNIVERSITY

Model Development – Generalized Regression Beta distribution

Response Distribution	Estimation Method	Validation Method	Nonzero Parameters	AICc	BIC	Generalized RSquare	Generalized RSquare
Beta	Maximum Likelihood	Validation Column	28	1889.4779	2036.396	-101.327	-34.040

Term	Estimate	Std Error	Wald ChiSquare	Prob > ChiSquare
Intercept	-3.863	0.127	932.562	<.0001
nor_ROA^2	0.367	0.287	1.634	0.201
nor_ROCE	1.373	0.139	98.225	<.0001
nor_ROA	-3.285	0.389	71.423	<.0001
nor_PreTaxPM	-0.238	0.115	4.261	0.039
SectorZ_ROCE*AT	-0.299	0.049	37.430	<.0001
SectorZ_DebtRatio	-0.140	0.034	17.253	<.0001
nor_EQmult^2	-0.057	0.055	1.059	0.303
nor_EVmult^2	0.016	0.001	200.673	<.0001
nor_DE_ratio^2	0.080	0.055	2.148	0.143
nor_FCFperShare^2	-0.001	0.003	0.183	0.669
nor_TaxRate^2	0.027	0.008	12.339	0.000
nor_DebtRatio^2	1.856	0.143	167.596	<.0001
FCF_OPCratio/DSO	-0.018	0.012	2.221	0.136
nor_CR*ROA	1.065	0.000	.	.
nor_OPM*DEratio	0.158	0.034	21.212	<.0001
nor_NOPAT_sin	-0.679	0.395	2.961	0.085
nor_EVmult_sin	-0.229	0.095	5.773	0.016
Adjusted_Nor_ROE	-1.015	0.145	49.035	<.0001
Composite_ProfitabilityScore	0.090	0.156	0.335	0.563
Nor_FCFperShare	-0.068	0.015	20.735	<.0001
nor_OPcPerShare	-0.136	0.017	67.462	<.0001
nor_OPc_Sales_Ratio	-0.078	0.064	1.499	0.221
nor_CurrentRatio	0.433	0.020	446.290	<.0001
nor_DE_ratio	-0.048	0.011	18.667	<.0001
nor_PayableTurn	0.001	0.017	0.001	0.975
nor_TaxRate	-0.088	0.026	11.348	0.001

Measure	Training	Validation
Number of rows	1462	367
Sum of Frequencies	1462	367
-LogLikelihood	916.172	-14.055
Number of Parameters	28	28
BIC	2036.396	137.240
AICc	1889.478	32.694
Generalized RSquare	-101.327	-34.040



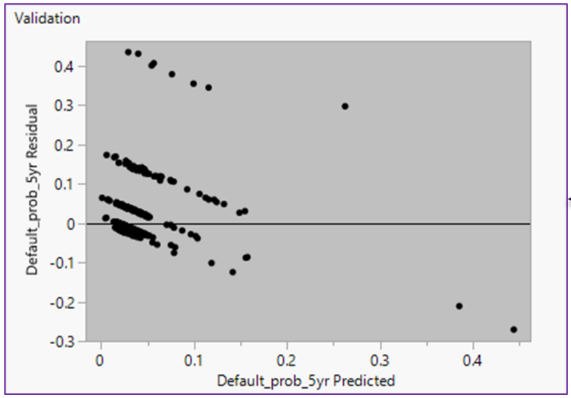
- But Rsquared values do not indicate this model predicts better than the Mean default probability

Model Development – Generalized Regression Beta distribution

Response Distribution	Estimation Method	Validation Method	Nonzero Parameters	AICc	BIC	Generalized RSquare	Generalized RSquare
Beta	Maximum Likelihood	Validation Column	28	1889.4779	2036.396	-101.327	-34.040

Term	Estimate	Std Error	Wald ChiSquare	Prob > ChiSquare
Intercept	-3.863	0.127	932.562	<.0001
nor_ROA^2	0.367	0.287	1.634	0.201
nor_ROCE	1.373	0.139	98.225	<.0001
nor_ROA	-3.285	0.389	71.423	<.0001
nor_PreTaxPM	-0.238	0.115	4.261	0.039
SectorZ_ROCE*AT	-0.299	0.049	37.430	<.0001
SectorZ_DebtRatio	-0.140	0.034	17.253	<.0001
nor_EQmult^2	-0.057	0.055	1.059	0.303
nor_EVmult^2	0.016	0.001	200.673	<.0001
nor_DE_ratio^2	0.080	0.055	2.148	0.143
nor_FCFperShare^2	-0.001	0.003	0.183	0.669
nor_TaxRate^2	0.027	0.008	12.339	0.000
nor_DebtRatio^2	1.856	0.143	167.596	<.0001
FCF_OPCratio/DSO	-0.018	0.012	2.221	0.136
nor_CR*ROA	1.065	0.000	.	.
nor_OPM*DEratio	0.158	0.034	21.212	<.0001
nor_NOPAT_sin	-0.679	0.395	2.961	0.085
nor_EVmult_sin	-0.229	0.095	5.773	0.016
Adjusted_Nor_ROE	-1.015	0.145	49.035	<.0001
Composite_ProfitabilityScore	0.090	0.156	0.335	0.563
Nor_FCFperShare	-0.068	0.015	20.735	<.0001
nor_OPcPerShare	-0.136	0.017	67.462	<.0001
nor_OPc_Sales_Ratio	-0.078	0.064	1.499	0.221
nor_CurrentRatio	0.433	0.020	446.290	<.0001
nor_DE_ratio	-0.048	0.011	18.667	<.0001
nor_PayableTurn	0.001	0.017	0.001	0.975
nor_TaxRate	-0.088	0.026	11.348	0.001

Measure	Training	Validation
Number of rows	1462	367
Sum of Frequencies	1462	367
-LogLikelihood	916.172	-14.055
Number of Parameters	28	28
BIC	2036.396	137.240
AICc	1889.478	32.694
Generalized RSquare	-101.327	-34.040



- Residual vs Predicteds still exhibit pattern from Log-linearized model… indicating poor fit

Model Development – Generalized (Penalized/Regularized) regression

	LASSO	Elastic Net	Ridge
Nonzero Parameters	27	27	28
AICc	🟩 -2994.389	-2994.381	🟥 -2992.494
BIC	🟩 -2852.680	-2852.671	🟥 -2845.576
Generalized RSquare	0.284108	🟥 0.284104	🟩 0.284198
Validation Generalized RSquare	🟩 0.203948	0.203941	🟥 0.203792
▨ Number of rows	367	367	367
▨ Sum of Frequencies	367	367	367
▨ -LogLikelihood	-421.6463	-421.6449	-421.6230
▨ Number of Parameters	27	27	28
▨ BIC	🟩 -683.8477	-683.8450	🟥 -677.8958
▨ AICc	🟩 -784.8323	-784.8296	🟥 -782.4412
▨ Generalized RSquare	🟩 0.203948	0.2039413	🟥 0.203792
▨ RASE	🟥 0.0757008	🟩 0.0757011	0.075708
Lambda Penalty	0.000832	0.0008407	0

▨ Results from model run on **Validation** data
🟩 "Best" Among models
🟥 "Worst" Among models

- Evaluating the Default Probability as a **continuous, normally distributed** variable yields similar results across the penalized regressions.

California Lutheran
UNIVERSITY

Model Development – Generalized (Penalized/Regularized) regression

	LASSO	Elastic Net	Ridge
Nonzero Parameters	27	27	28
AICc	🟩 -2994.389	-2994.381	🟥 -2992.494
BIC	🟩 -2852.680	-2852.671	🟥 -2845.576
Generalized RSquare	0.284108	🟥 0.284104	🟩 0.284198
Validation Generalized RSquare	🟩 0.203948	0.203941	🟥 0.203792
Number of rows	367	367	367
Sum of Frequencies	367	367	367
-LogLikelihood	-421.6463	-421.6449	-421.6230
Number of Parameters	27	27	28
BIC	🟩 -683.8477	-683.8450	🟥 -677.8958
AICc	🟩 -784.8323	-784.8296	🟥 -782.4412
Generalized RSquare	🟩 0.203948	0.2039413	🟥 0.203792
RASE	🟥 0.0757008	🟩 0.0757011	0.075708
Lambda Penalty	0.000832	0.0008407	0

Results from model run on **Validation** data
🟩 "Best" Among models
🟥 "Worst" Among models

- Evaluating the Default Probability as **a continuous, normally distributed** variable yields similar results across the penalized regressions.
- *Note:* Statistics present in the table are using Validation Method of **Validation Column**

Kfold validation	LASSO	Elastic Net	Ridge
NonZero Parameters	21	23	27

Model Selection – How to they compare

Model	R^2	SSE	RASE
OLS	0.20380	10.63079	0.07571
PLS	0.28138	12.74948	0.08349
KNN	0.22597	2.35068	0.07609
Generalized Linear	-34.04000	13.40371	0.08561
LASSO	0.20395		0.07570
Elastic Net	0.20394		0.07570
Ridge	0.20379		0.07571

*RASE for GLM and PLS computed using “Predicteds”
*GLM is from Beta Distribution

*SSE approximated For Penalized regression using calculation:

$$SSE = \left(scale\ Estimate * sqrt(n)\right)^2$$

*SSE for penalized models was difficult to obtain

🟩 "Best" Among models
🟥 "Worst" Among models

Model Selection – How to they compare

Model	R^2	SSE	RASE
OLS	0.20380	10.63079	0.07571
PLS	0.28138	12.74948	0.08349
KNN	0.22597	2.35068	0.07609
Generalized Linear	-34.04000	13.40371	0.08561
LASSO	0.20395		0.07570
Elastic Net	0.20394		0.07570
Ridge	0.20379		0.07571

*RASE for GLM and PLS computed using "Predicteds"
*GLM is from Beta Distribution

*SSE approximated For Penalized regression using calculation:

$$SSE = \left(scale\ Estimate * sqrt(n)\right)^2$$

*SSE for penalized models was difficult to obtain

🟩 "Best" Among models
🟥 "Worst" Among models

Model Ranking on Favorable Metric (RASE)

1. LASSO
2. Elastic Net
3. OLS
4. RIDGE
5. KNN
6. GLM

Model Selection – How to they compare

Model	R^2	SSE	RASE
OLS	0.20380	10.63079	0.07571
PLS	🟩 0.28138	12.74948	0.08349
KNN	0.22597	🟩 2.35068	0.07609
Generalized Linear	-34.04000	13.40371	0.08561
LASSO	0.20395		🟩 0.07570
Elastic Net	0.20394		🟩 0.07570
Ridge	0.20379		0.07571

*RASE for GLM and PLS computed using "Predicteds"
*GLM is from Beta Distribution

*SSE approximated For Penalized regression using calculation:

$SSE = (scale\ Estimate * sqrt(n))^2$

*SSE for penalized models was difficult to obtain

🟩 "Best" Among models
🟥 "Worst" Among models

Model Ranking on Favorable Metric (RASE)

1. LASSO
2. Elastic Net
3. OLS
4. RIDGE

^These 4 are very close… One final evaluation

Final Evaluation – What if we model Def_prob_5yr as BINARY outcome

WHY?

- A “Default event” is inherently binary (0 or 1), a company either defaults or they do not
- We can gain additional Predictive performance measures from this analysis

Final Evaluation – What if we model Def_prob_5yr as BINARY outcome

WHY?

- A “Default event” is inherently binary (0 or 1), a company either defaults or they do not
- We can gain additional Predictive performance measures from this analysis

Context:

- New Y variable “Def_Prob_Bin2”
- Rule: IF *Def_Prob_5yr* > .06 then 1 else 0
- Variable creates 864 *default* cases (1) and 1165 NO default (0) cases in data set

Model Selection– Think about the cost of missing a default

Elastic Net wins:

- Default is the positive class -- Sensitivity metric is given priority
- Elastic Net outperforms on overall accuracy in correctly predicting True Negatives (Non-Defaults) too

Elastic Net

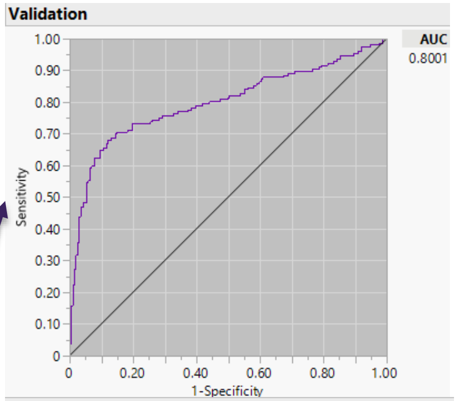
Method	TP	FN	FP	TN	Sensitivity	Specificity	Precision	Accuracy	F1	MCC
Fit Generalized	99	58	20	189	0.6306	0.9043	0.8319	0.7869	0.7174	0.5651

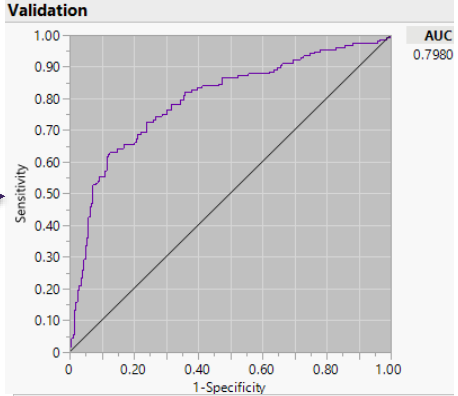
LASSO

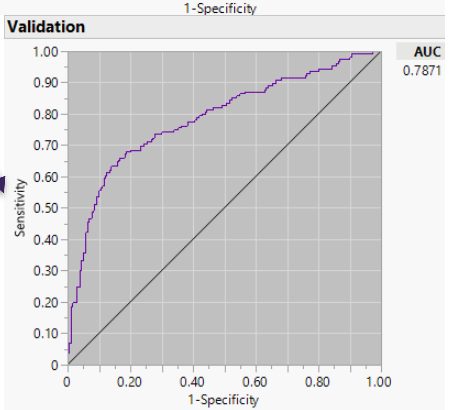
Method	TP	FN	FP	TN	Sensitivity	Specificity	Precision	Accuracy	F1	MCC
Fit Generalized	88	68	22	188	0.5641	0.8952	0.8	0.7541	0.6617	0.4954

Ridge

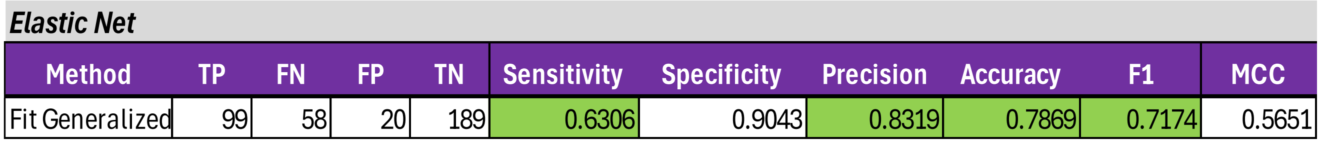
Method	TP	FN	FP	TN	Sensitivity	Specificity	Precision	Accuracy	F1	MCC
Fit Generalized	99	58	28	181	0.6306	0.866	0.7795	0.765	0.6972	0.5163





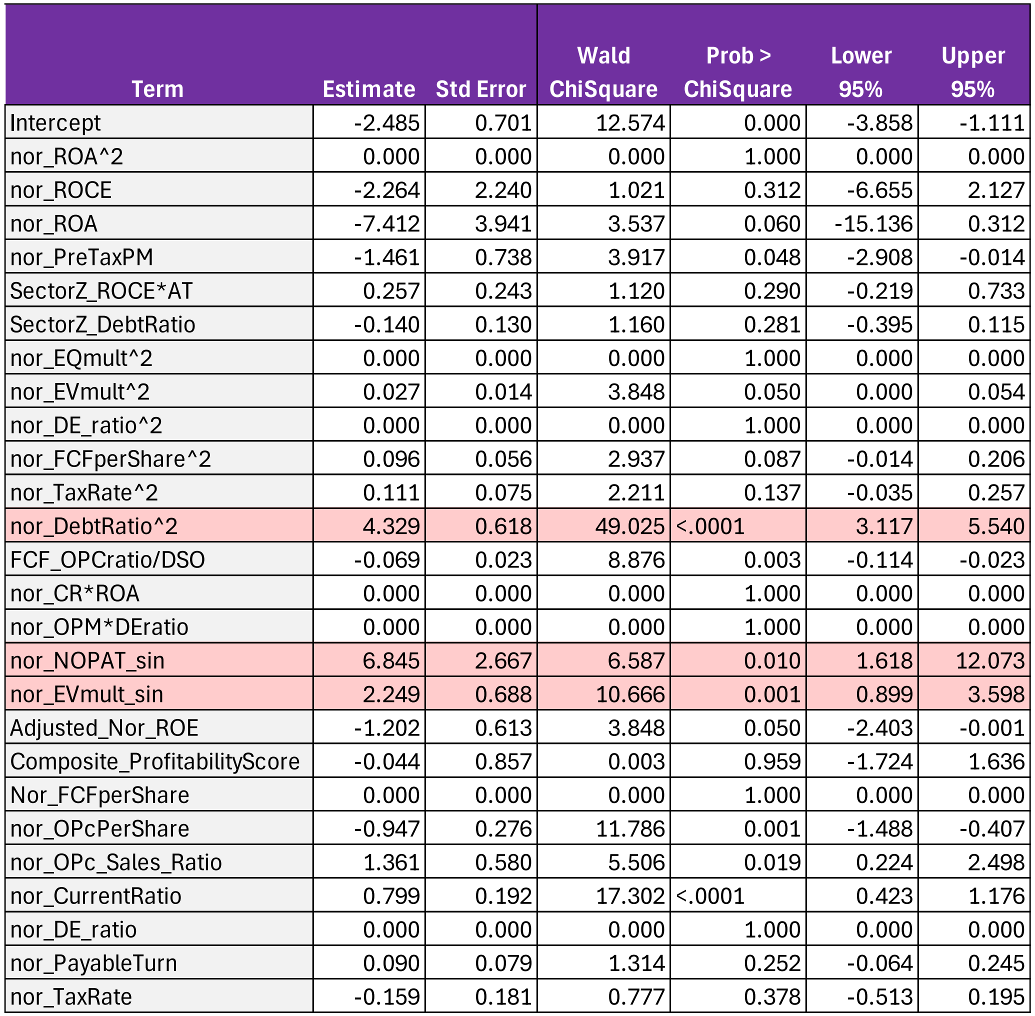


Conclusion- Elastic Net is the winning Model

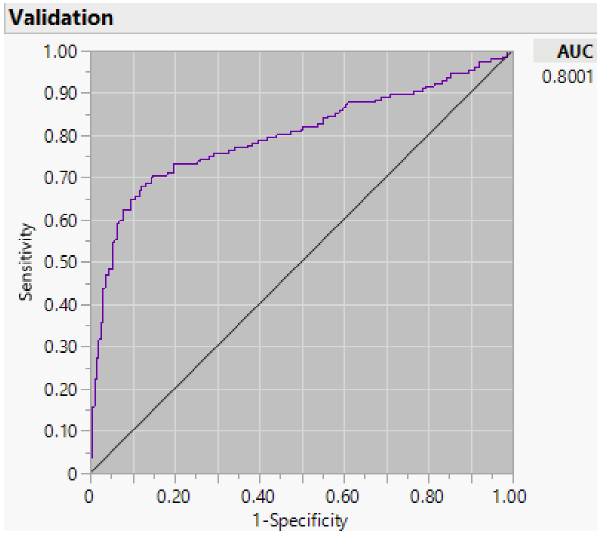


Elastic Net

Method	TP	FN	FP	TN	Sensitivity	Specificity	Precision	Accuracy	F1	MCC
Fit Generalized	99	58	20	189	0.6306	0.9043	0.8319	0.7869	0.7174	0.5651



Term	Estimate	Std Error	Wald ChiSquare	Prob > ChiSquare	Lower 95%	Upper 95%
Intercept	-2.485	0.701	12.574	0.000	-3.858	-1.111
nor_ROA^2	0.000	0.000	0.000	1.000	0.000	0.000
nor_ROCE	-2.264	2.240	1.021	0.312	-6.655	2.127
nor_ROA	-7.412	3.941	3.537	0.060	-15.136	0.312
nor_PreTaxPM	-1.461	0.738	3.917	0.048	-2.908	-0.014
SectorZ_ROCE*AT	0.257	0.243	1.120	0.290	-0.219	0.733
SectorZ_DebtRatio	-0.140	0.130	1.160	0.281	-0.395	0.115
nor_EQmult^2	0.000	0.000	0.000	1.000	0.000	0.000
nor_EVmult^2	0.027	0.014	3.848	0.050	0.000	0.054
nor_DE_ratio^2	0.000	0.000	0.000	1.000	0.000	0.000
nor_FCFperShare^2	0.096	0.056	2.937	0.087	-0.014	0.206
nor_TaxRate^2	0.111	0.075	2.211	0.137	-0.035	0.257
nor_DebtRatio^2	4.329	0.618	49.025	<.0001	3.117	5.540
FCF_OPCratio/DSO	-0.069	0.023	8.876	0.003	-0.114	-0.023
nor_CR*ROA	0.000	0.000	0.000	1.000	0.000	0.000
nor_OPM*DEratio	0.000	0.000	0.000	1.000	0.000	0.000
nor_NOPAT_sin	6.845	2.667	6.587	0.010	1.618	12.073
nor_EVmult_sin	2.249	0.688	10.666	0.001	0.899	3.598
Adjusted_Nor_ROE	-1.202	0.613	3.848	0.050	-2.403	-0.001
Composite_ProfitabilityScore	-0.044	0.857	0.003	0.959	-1.724	1.636
Nor_FCFperShare	0.000	0.000	0.000	1.000	0.000	0.000
nor_OPcPerShare	-0.947	0.276	11.786	0.001	-1.488	-0.407
nor_OPc_Sales_Ratio	1.361	0.580	5.506	0.019	0.224	2.498
nor_CurrentRatio	0.799	0.192	17.302	<.0001	0.423	1.176
nor_DE_ratio	0.000	0.000	0.000	1.000	0.000	0.000
nor_PayableTurn	0.090	0.079	1.314	0.252	-0.064	0.245
nor_TaxRate	-0.159	0.181	0.777	0.378	-0.513	0.195


Validation
AUC
Sensitivity
1-Specificity