RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

A Predictive Ordered Logistic Regression Model as a Tool for Quality Review of Control Risk Assessments

Background

- Management and external auditors are required to report on the adequacy of internal controls *(SOX 404)*
- Internal audit quality is important to external auditors as well as management *(Gramling & Vandervelde, 2006)*
- External auditors are encouraged to take the work of internal auditors into consideration (AS No. 5)
- SOX requires external auditors to report on the adequacy of the internal controls as well as the management's assessment
- Control Risk Assessments (CRA): a popular tool that helps the auditors to get a better understanding of business processes

Motivation, Research Questions, & Findings

Motivation

- Need to develop a methodology for the evaluation of CRA by internal auditors and CRSA by business owners (quality review)
- Need to prioritize identified exceptions (cases that deviate from the predicted values)

Research Questions

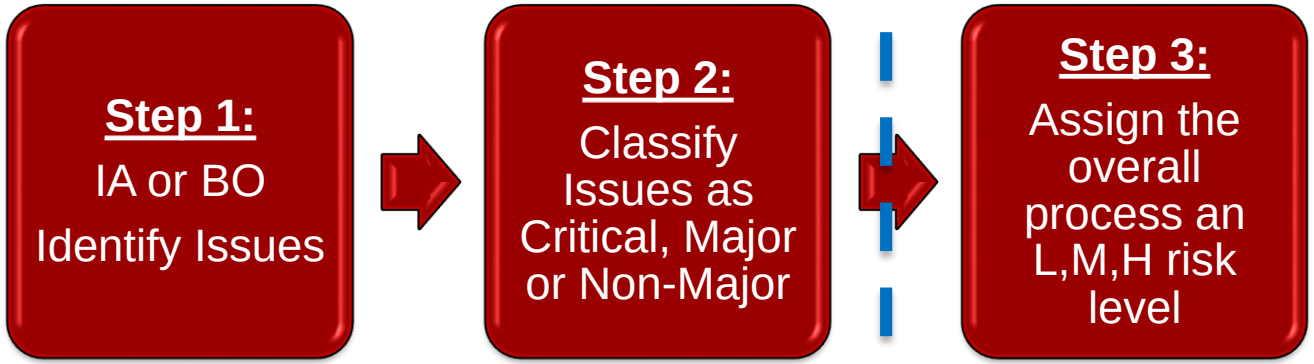
1. How can we verify and review the quality of internal auditors' (business owners') judgment in control risk assessments?
2. How can we prioritize the exceptions that deviate from the norms?

Findings

- CRA: accuracy of fitted model is 83%, predictive model 76.36%
- CRSA: accuracy of fitted model 74.32%, predictive model 76.5%
- Business owners tended to overestimate risk, but showed signs of improvement with time (gained experience)
- Feedback from the company indicates that the ranking metrics were effective

Data

- Source: Multinational consumer products company
- Issues identified by location and business process (e.g. Distribution, Payroll, Purchasing, A/P)



- Data breakdown:

	FY 08/09	FY 09/10	FY 10/11	All (08-11)
CRA	344	305	275	924
CRSA	3310	3138	3145	9593

Ordered Logistic Regression

- Variables: ordinal and labeled (audit risk levels)
- **Ordered Logistic Regression:**

$$y_i^* = logit = ln\left(\frac{prob(event)}{1 - prob(event)}\right) = \beta^T x_i + \varepsilon_i = \beta_0 + \beta_1 CC + \beta_2 MC + \beta_3 NMC$$

- y_i^*= **logit** = log of the odds that a certain event takes place.
- β_0 = Intercept
- β_i = Coefficient
- **CC** = Number of critical issues (identified by auditors/business owners)
- **MC** = Number of Major issues (identified by auditors/business owners)
- **NMC** = Number of Non-Major issues (identified by auditors/business owners)

- $y_i = 0 \quad for \quad y_i^* \leq \mu_L \quad (Low\ risk)$
- $y_i = 1 \quad for \quad \mu_L < y_i^* \leq \mu_M \quad (Medium\ risk)$
- $y_i = 2 \quad for \quad \mu_M < y_i^* \leq \mu_H \quad (High\ risk)$

Predicted Probabilities Calculations

Predicted probability:

- $PredProb = \widehat{P}(C_i|x) = \frac{1}{1+e^{-(\beta^T x_i+\varepsilon_i)}}$
  - β^T is a vector of Intercepts
  - x_i is the vector of coefficients
  - The class with the highest calculated probability is the predicted class

- $Calc\_H = \frac{1}{1+e^{-\{(Intercept\_2+(CC\_Coeff*CC)+(MC\_Coeff*MC)+(NMC\_Coeff*NMC)\}}}$
- $Calc\_M = \left(\frac{1}{1+e^{-\{(Intercept\_1+(CC\_Coeff*CC)+(MC\_Coeff*MC)+(NMC\_Coeff*NMC)\}}}\right) - Calc\_H$
- $Calc\_L = 1 - Calc\_H - Clac\_M$

Outliers Identification and Ranking

Record	CC	MC	NMC	Calc_H	Calc_M	Calc_L	Assign. Class	Pred. Class	Ratio	Diff.
123456	0	2	3	0.60719	0.39195	0.00086	M	H	**0.64551**	**0.21524**

Outliers' disagreement measure:

$$Ratio = \frac{Calc.prob\_Assigned\ Class}{Calc.prob\_Predicted\ Class}$$

$$Difference = Calc.prob\_Predicted\ Class - Calc.prob\_Assigned\ Class$$

$$Ratio = \frac{0.39195}{0.60719} = 0.64551$$

$$Difference = 0.60719 - 0.39195 = 0.21524$$

The lower (bigger) the ratio (difference), the more suspicious the record is

Findings – Auditors

- Accuracy: 83% (fitted model), 76.36% (predictive model)
- Sliding window technique shows consistency of results

Confusion Matrix-Fitted Model (M08/10-D08/10)				
Predicted Level	Assigned Level			
	L	M	H	Total
L	327 88.38%	43 11.62%	0 0.00%	370
M	40 16.46%	184 75.72%	19 7.82%	243
H	0 0.00%	8 22.22%	28 77.78%	36
Total	367	235	47	649

Confusion Matrix-Predictive Model (M08/10-D10/11)				
Predicted Level	Assigned Level			
	L	M	H	Total
L	142 83.53%	25 14.71%	3 1.76%	170
M	23 27.05%	58 68.24%	4 4.71%	85
H	0 0.00%	10 50%	10 50%	20
Total	165	93	17	275

Findings – Business Owners

- Accuracy: 74.32% (fitted model), 76.5% (predictive model)

Confusion Matrix-Fitted Model (M08/10-D08/10)				
Predicted Level	Assigned Level			
	L	M	H	Total
L	3817 74.38%	1298 25.29%	17 0.33%	5132
M	200 16.43%	930 76.42%	87 7.15%	1217
H	1 1.01%	53 53.54%	45 45.45%	99
Total	4018	2281	149	6448

Confusion Matrix-Predictive Model (M08/10-D10/11)				
Predicted Level	Assigned Level			
	L	M	H	Total
L	1822 77.76%	519 22.15%	2 0.09%	2343
M	119 16.46%	554 76.63%	50 6.92%	723
H	3 3.80%	46 58.23%	30 37.97%	79
Total	1944	1119	82	3145

Other Findings

Auditors vs. Business Owners:

- Extreme Outliers in Fitted model: BO 18 – IA 0
- Extreme Outliers in Predictive model: BO 5 – IA 3
- Highest level of Disagreement: Pred. H – Assigned M (BO & IA)
  - Reluctance to assign high risk levels due to the possible ramifications
- Predictive Model Accuracy: increased for BO, decreased for IA
  - Started using CRSA in 2008 – BO gained experience with time

Interesting finding:

- 3 records with no issues, but High risk
- Systematic bias to overestimate risk level by BO in general
  - Conservatism

Conclusion

Contribution:

- Proposed a methodology to review the quality of auditors'/Business Owners' assessments of control risk
- Proposed a methodology to prioritize outliers, thus increasing audit efficiency by helping auditors focus their efforts on more suspicious records

Limitations:

- Distance between variables is unknown (L-M vs. M-L and NM-M vs M-C etc)
- Unbalanced datasets (although this is the real life scenario)
- Unknown issues categorization criteria (by the company)

Future Research:

- Develop more sophisticated ranking techniques and compare their performance
- Use bigger datasets (more years)