

Credit Scoring and the Edinburgh Conferences: Fringe or International Festival

Lyn Thomas

University of Southampton

August 2011

12th International Conference on Credit Scoring and Credit Control

UNIVERSITY OF
Southampton
School of Management

- Credit Scoring and Credit Control conferences began in the University of Edinburgh in 1989
 - Run every other year since then
- My privilege to suggest the idea soon after arriving at Edinburgh University
 - Objective was to have a “level playing field” to discuss developments in consumer credit risk assessment
 - Make links between researchers and practitioners in lending organisations and research institutions
 - Encourage research on consumer credit risk modelling in universities
- My pleasure together with Jonathan Crook and David Edelman to have organised all subsequent conferences
- Similar conferences started in Canada (1995) and Australia (2011)₂



- Established in 1947
 - World class cultural events reflecting international culture
 - Recognised as one of the main showcases for the best of the best
 - Initially could not have conference at same time as festival because University wanted all accommodation for visitors
 - In 2005, University went to semesters starting at beginning of September so agreed to give conference accommodation even if festivals running
- Also started in 1947 (given its name in 1948) when theatre companies turned up for other festival uninvited
 - Fringe is unjuried, seeks original and sometime controversial acts

Objectives

- Conference has given window on developments in consumer risk assessment in last 25 years. What are these developments in the last 25 years?
- What has happened to the ideas suggested in the conference?
 - Which have become world class (part of the international festival)
 - Which are still being trialled ?(part of the fringe)
 - Which have not been pursued?
- What are the current challenges to consumer credit risk modelling?

Four Ages of the Conference I

- 1989–1993: Stabilising scorecard building tools
 - Behavioural scoring becoming established
 - Multiple scorecard building approaches tested
 - Survival analysis (now becoming international)
 - Expert systems (now used in fraud)
 - Neural nets
 - Bayesian learning
 - Flat maximum effect
 - Reject inference
 - augmentation/extrapolation depend on assumptions
use credit bureau info or sampling below cut off
 - Lead taken by retailers, consultancies (not banks)
 - Dividing credit card costs fairly

Four Ages of the Conference II

- 1995–2001: Profitable decision making
 - Business measures of scorecard performance
 - Risk based pricing
 - Markov chain models for credit limits
 - Experimental design
 - Small business/mortgage modelling
 - 1st paper on macro economic changes affect scorecard
 - Affordability started to be discussed
 - New methods : nearest neighbour, graphical methods
 - Banks and consultancies to the fore
 - Mainly UK/US with some Western European focus

Four Ages of the Conference III

- 2003–2007: Basel and banking
 - Basel Accord
 - Basel first mentioned in my talk in 2001
 - Next three conferences dominated by Basel regulations
 - Regulators spoke at conferences
 - Validation of scorecards, Low default portfolios
 - Other themes continue
 - Optimal pricing of loans, adverse selection
 - Customer lifetime value
 - Data guided decision making
 - New methods
 - Genetic algorithms, Quantile regression (fringe?)
 - 2007: financial crisis occurs but only 3 related papers
 - subprime mortgage scoring, choice of fixed or variable rate mortgages; distressed debt (saying things were getting better)
 - Real international focus with papers from Eastern Europe, Australia, Brazil, Korea

Four Ages of the Conference IV

- 2009–2011: Dynamics and Debt
 - Dynamics of scorecard movement
 - Macro economic impact on score
 - Dynamics of the score – long run average PD; stress testing in Basel
 - Markov chain models
 - Systemic risks and modelling portfolio level credit risk
 - Stress testing
 - Dual time dynamics
 - Loss Given Default and Collections Modelling
 - LGD models needed in Basel
 - No good models of how collection actions affects LGD
 - Modelling borrowers motivation not just performance
 - Psychometrics
 - Use of social networks
 - Impact of foreclosures
 - Causal effect of credit decisions

The “international festival” themes and their current challenges

- Methodologies for credit scoring
- Making profitable decisions
- The dynamics of scores and introducing economics into them
- Portfolio level consumer credit risk models
- Measuring how good are scorecards
- Debt and the collections process

1: Methodologies for scoring

- Established:
 - logistic regression (90%+ built this way)
 - Linear regression, classification trees , LP (for certain circumstances)
- Becoming established:
 - Survival analysis (suggested in 89, 30+ papers since then)
- Possible Future “International” Approaches
 - Quantile regression
 - Bayesian modelling

Methodologies

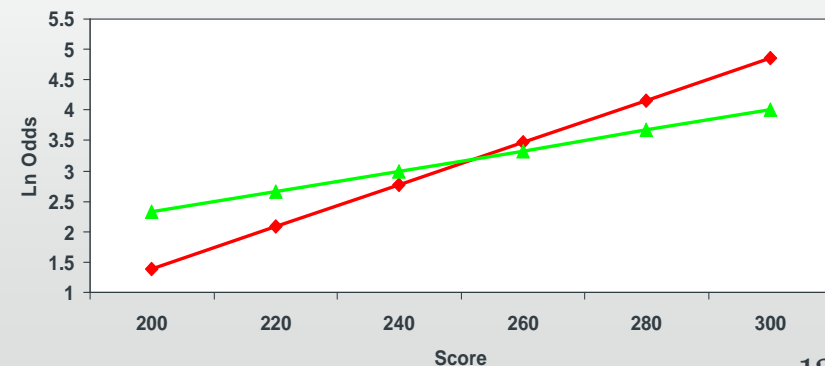
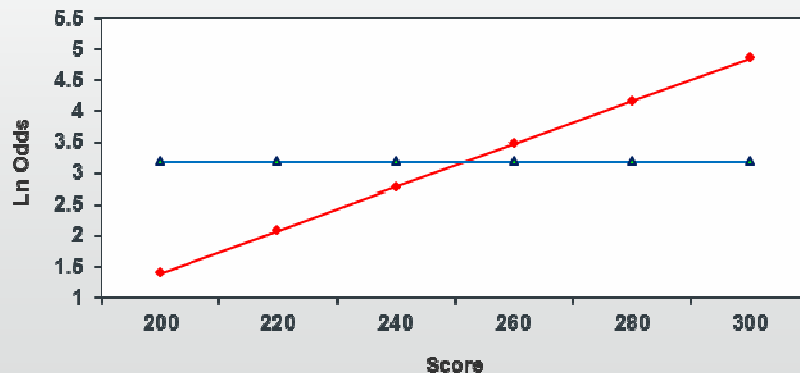
- Fringe
 - Support Vector machines – good results but black box
 - Neural Nets– established in other data mining areas
 - Ensemble Methods – good results but can they be “marketed” internally and externally
 - Nearest Neighbours– updates scorecard quickly
 - Linear and integer programming – size of problem
 - Genetic algorithms
 - Bayesian learning
 - Expert systems (1989 but now in fraud detection)
- Comparison methodology
 - Use real sized data sets (not German 1000 case example)
 - Variable modification/ binning should be independent of methodology
 - Choice of variables methodology dependent
 - Not just one outcome, but cross validation and whether differences significant

2. Making profitable decisions

- Application Process
 - Variable pricing (international)
 - Adverse selection in variable pricing
 - Price of other features than interest rate
 - Need to estimate take (in cards this is use) probability
 - » Experimental design (fringe topic)
 - Joint propensity/default risk/ product choice modelling
- Operational Process
 - Behavioural scoring for credit limit decisions (international)
 - Markov chains and Markov decision processes to include dynamics
 - Customer lifetime value (fringe); customer level decisions.
Propensity to cross/up sell
- Collections Process (future)
- Need action/response modelling
 - How to include lenders' action in borrowers' performance

3. What is the dynamics of a score

- Surprisingly little modelling of how an individual's score changes over time
- Original view was score should be static so movement considered as “failure” of score.
- Log odds – score graphs suggest how scorecard is changing but not individual scores



What is the dynamics of a score

- As a diffusion process is it
 - Random walk (De Andrade suggest this for bureau data)
 - Has trend and seasonality ?
 - Jump process ?
- If split score into bands get discrete space
 - So a Markov chain but what order
 - Malik, Scherer suggest second order better than first
 - Non stationary
 - depending on economics, time on books
- Simple decomposition makes clear how economic may affect score

Dynamics and economics into credit scores

- Given characteristics \mathbf{x} of borrower/loan, score is sufficient statistic for probability of being Good

$$\Pr\{\text{Good} \mid \text{score based on } \mathbf{x}\} = p(G \mid s(\mathbf{x})) = p(G \mid s(\mathbf{x}), \mathbf{x}) = p(G \mid \mathbf{x}) \quad \forall \mathbf{x} \in \mathcal{X}$$

- Log odds score is

$$s(x) = \log\left(\frac{p(G \mid x)}{p(B \mid x)}\right)$$

$$p(G \mid \mathbf{x}) = \frac{e^{s(\mathbf{x})}}{1 + e^{s(\mathbf{x})}} = \frac{1}{1 + e^{-s(\mathbf{x})}}$$

- Bayes theorem splits score into population odds plus weights of evidence (impact of individual characteristics) ; if p_G (p_B) proportions of Goods (Bads) in population

$$s(\mathbf{x}) = \ln\left(\frac{p_G p(\mathbf{x} \mid G)}{p_B p(\mathbf{x} \mid B)}\right) = \ln\left(\frac{p_G}{p_B}\right) + \ln\left(\frac{p(\mathbf{x} \mid G)}{p(\mathbf{x} \mid B)}\right) = \ln o_{pop} + \ln I(\mathbf{x}) = s_{Pop} + s_{Inf}(\mathbf{x})$$

Why introduce economic and market variables into score?

- Normally scores thought of as static but they are really dynamic : want score at time t to be

$$s(\mathbf{x}, t) = s_{Pop}(t) + woe(\mathbf{x}, t)$$

- What scorecard gives is $s(\mathbf{x}, t_0) = s_{Pop}(t_0) + woe(\mathbf{x}, t_0)$ where t_0 is when scorecard built

- Solution: put economic conditions, $\mathbf{e}(t)$, into scorecard

$$s(\mathbf{x}, t) = s(\mathbf{x}, \mathbf{e}(t)) = s_{Pop}(\mathbf{e}(t)) + woe(\mathbf{x}, \mathbf{e}(t))$$

- Obviously $s_{pop}(\mathbf{e})$ depends on \mathbf{e}

- $s_{pop}(\mathbf{e})$ is transformation of population default rate; must change over time

$$s_{pop}(\mathbf{e}(t)) = c_1 e_1(t) + \dots + c_m e_m(t)$$

- Does $woe(\mathbf{x}, \mathbf{e})$ depend on \mathbf{e} : if so need interaction terms between economic variables and borrower characteristics

$$woe(\mathbf{x}, t) = \sum_{i=1}^m \sum_{j=1}^n c_{ij} (IndicatorFunction(x_j) e_i(t))$$

- Which economic variables to use ? Will be different for different products¹⁶

Case Study: Invoice Discounting Example

- Scorecard built to estimate default risk of small firms, where bank is invoice discounter (like factoring)
- Give loan using firm's invoices to customers as collateral
- Scorecard built circa 2005/6 continued to discriminate well through 2009
- But estimate of number of firm's defaulting grossly underestimated in 2008/9.

Scorecard without Economic Variables

	Training	In time Test	Out of time Test
Gini	62	63	60
KS	46.46	48.83	46.34
HLtest (Chi square)	43.16	26.93	1470.94
Actual defaults	4666	2247	1409
Expected defaults	4666	2201	605

Scorecard with Confidence Index' and 'FTSE' as economics estimate $s_{pop}(e)$

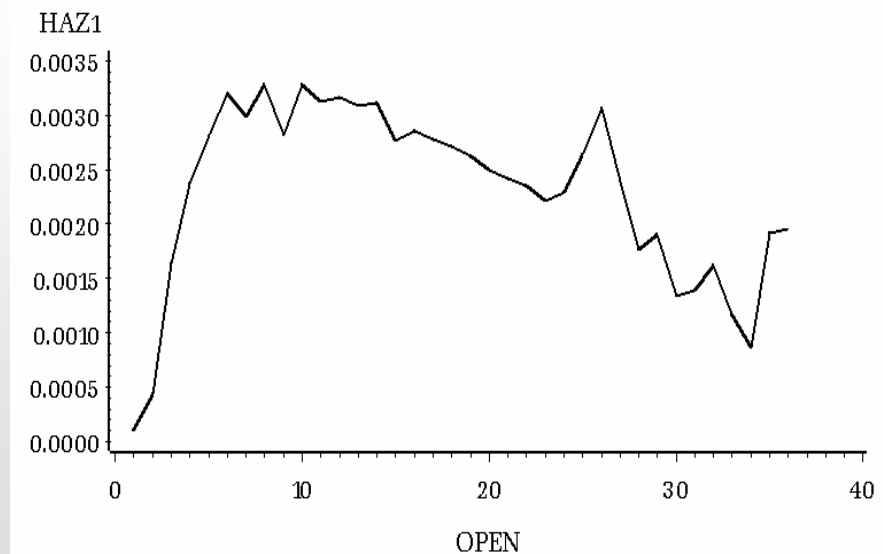
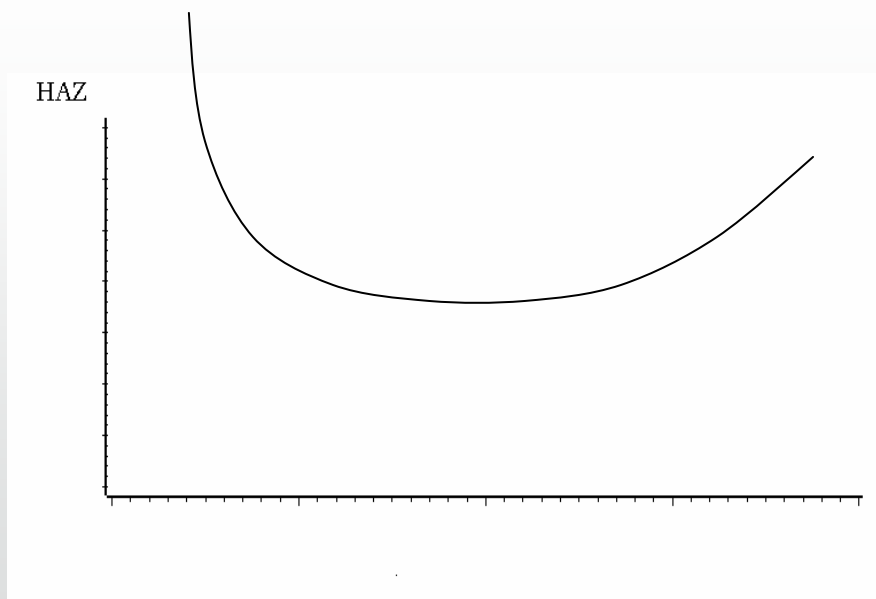
Version 1	Training	In time Test	Out of time Test
Gini	63	63	59
KS	47.12	49.00	48.80
HLtest (Chi square)	32.51	29.95	63.81
Actual defaults	4666	2247	1409
Expected defaults	4666	2202	1306

Model with interactions (Confidence and FTSE): economics estimate of $s_{pop}(e)$ and $woe(e)$

	Training	In time Test	Out of time Test
Gini	63	63	57
KS	47.03	49.27	44.17
HL(Chi square)	31.75	21.49	81.69
Actual defaults	4666	2247	1409
Expected defaults	4666	2204	1383

Including economic effects into hazard scorecards

- Standard scorecards (Logistic regression) have fixed outcome period; estimate chance of default at end of that period
- Can use survival analysis ideas to estimate when default T will occur depending on age of loan
- Use hazard function $h(t) = P(T = t | T \geq t)$



Proportional Hazard model

- Proportional Hazard models
 - Hazard function for default at time t for borrower with characteristics \mathbf{x} is

$$h(t, \mathbf{x}) = h_0(t) e^{-s(\mathbf{x})}$$

where $s(\mathbf{x})$ is hazard score

$$s(\mathbf{x}) = c_1 x_1 + \dots + c_n x_n$$

$$s = -\log(-\log p) \text{ (Hazard function)} : s = \log\left(\frac{p}{1-p}\right) \text{ (logistic regression)}$$

- Resulting scorecard is very similar to logistic regression

Extend to economic variables

- If t is time since loan started (started at t_0), the hazard of default at time $t+k$ for a person \mathbf{x} with economic conditions $EcoVar(k+t+t_0)$ and hazard score $s(\mathbf{x})$ is

$$h(t+k, \mathbf{x}) = h_0(k+t) e^{(-\alpha s(\mathbf{x}) + \beta EcoVar_i(k+t+t_0))}$$

Months on
Books Factor

Idiosyncratic
Risk

Systemic
Risk

- Could replace $EcoVar(k+t+t_0)$ with $EcoVar(t+t_0)$ and so only use current economic variable values rather than forward forecasts
- Can redefine score so that

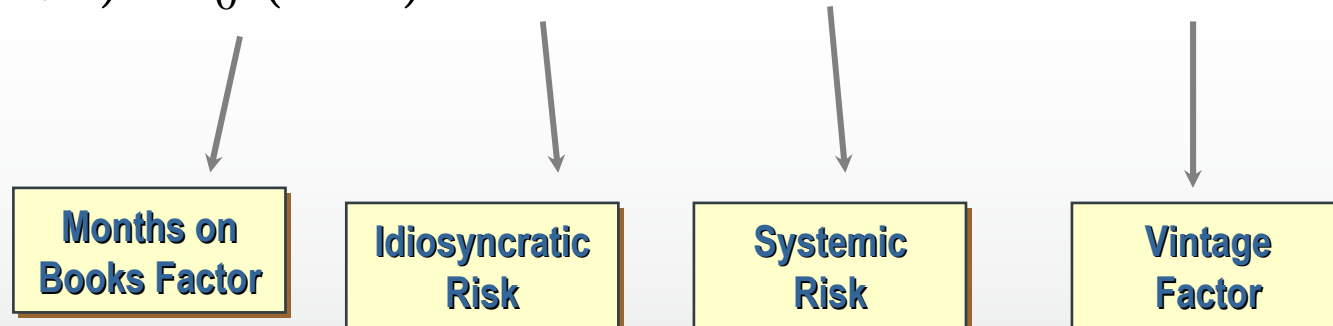
$$S(x, t) = -\alpha s(x) + \beta EconVar(t + t_0 + k)$$

- See Belloti and Crook, JORS; Malik and Thomas IJF)

Vintage Effect: economics does not explain everything

- As it stands all time dependent effects are to be explained by “economic variables”.
- In reality some are changes in policy
- Can explain this by including “vintage effect”

$$h_i(t+k, \mathbf{x}) = h_0(k+t) e^{(-\alpha s(\mathbf{x}) + \beta EcoVar_i(k+t+t_0) + \gamma Vintage(t_0))}$$



- At segment level this is very similar to dual time dynamics approach

Use Macro Economics in score to PD transformation

- Instead of including economic effects in scorecard , leave scorecard fixed and include economics in score to PD transformation
- Can reinterpret hazard model in this way
- If t is time since loan started (started at t_0), the hazard of default at time $t+k$ for a person i with economic conditions $EcoVar(k+t+t_0)$ and hazard score $s(\mathbf{x})$ is

$$h(t+k, \mathbf{x}) = h_0(k+t) e^{(-\alpha s(\mathbf{x}) + \beta EcoVar_i(k+t+t_0))}$$

Score:
 Idiosyncratic
 Risk

Systemic Risk
 transforming score
 to PD

- So now EconVar does not change the score just the transformation

$$s(x) \xrightarrow{EconVar} h(t, x) \rightarrow PD(t)$$

4. Portfolio Level credit risk models for consumer loans

- 50 years of credit risk modelling of individual loans
- Yet until recently no models for credit risk of portfolios of such loans
 - Basel II used corporate credit risk models
 - Mis rating of RMBS by rating agencies in 2006–8
- Essential to develop such models
- Currently Three approaches
 - Bottom Up 1 : use individual loan model (scores) and calculate default correlations
 - Bottom Up 2: use economic variables in individual loan models. These give default correlations
 - Top Down : start at segment level and use characteristics of loan that makes up the segment

Bottom Up 1

- Too many loans to get individual correlations so either
 - Common correlation (or common correlation with factor as in Basel)
 - Split into segments with correlations between segments (vintage, loan characteristics. Score)
 - Copula ideas – though not as popular in finance as they were
- Should correlations change over time because of the economics?

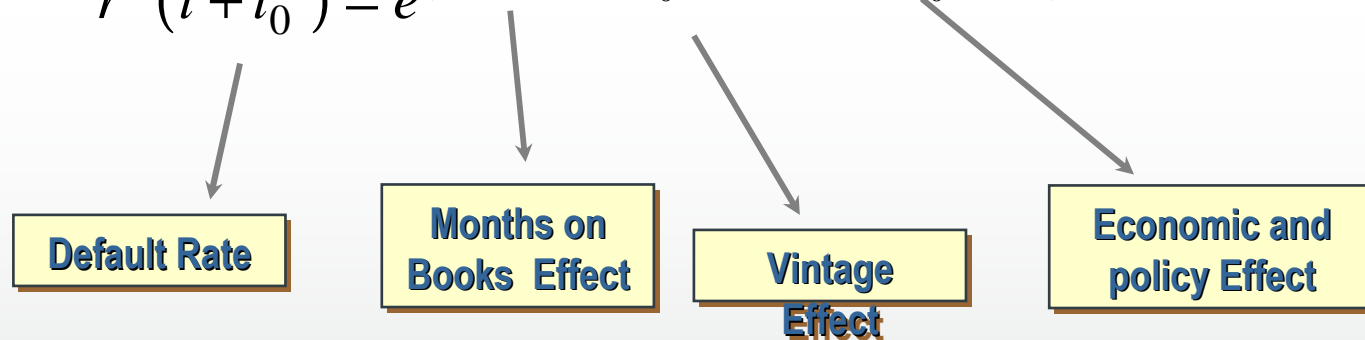
Bottom Up 2

- Earlier suggested extending scoring models for individual loans to include
 - Performance (Behavioural score)– so individual loan risk will be regularly updated
 - Economic and Market Variables– which ones are suitable depends on asset type
 - Origination Quality– one way of dealing with changes in lender's policy and with changes in borrower's market
 - Age of Loan – recognize the months on books effect
- Now use fact same economic and market variables on all loans to give correlations.
- Types of models do far developed with economic effects
 - Structural Models based on reputation
 - Hazard Models
 - Markov Chain Models

Top Down

- One example is Dual Time Dynamics (Breeden)
- Estimate at portfolio or at least segment level, “rate” of default $r(\cdot)$.
- For loans which started at time t_0 and current months on books is t , and economic variables are $e(t+t_0)$ then

$$r(t+t_0) = e^{(m(t)+v(t_0)+g(e(t+t_0))+\epsilon)}$$



- This also uses maturity, vintage, and economics as the factors that affect default.
- Difference is that estimates initially made at segment level
- Recent developments have integrated this with individual loan level

Measuring how good scorecard are

- Three different aspects of a default scorecard's performance that one might want to measure:

1. Discriminatory power (only uses scorecard)

- How good is the system at separating the goods and bads
 - Divergence - difference in expectations of weights of evidence
 - Kolmogorov Smirnov- difference in distribution function
 - ROC curves- comparison of distribution function
 - Gini coefficient/D concordance statistic

2. Calibration of probability forecast (uses scorecard and population odds)

- Not used much until Basel requirements and so few tests
 - Chi-square (Hosmer-Lemeshow) test
 - Binomial and Normal tests

3. Categorical prediction error (uses scorecard, population odds and cut-off score)

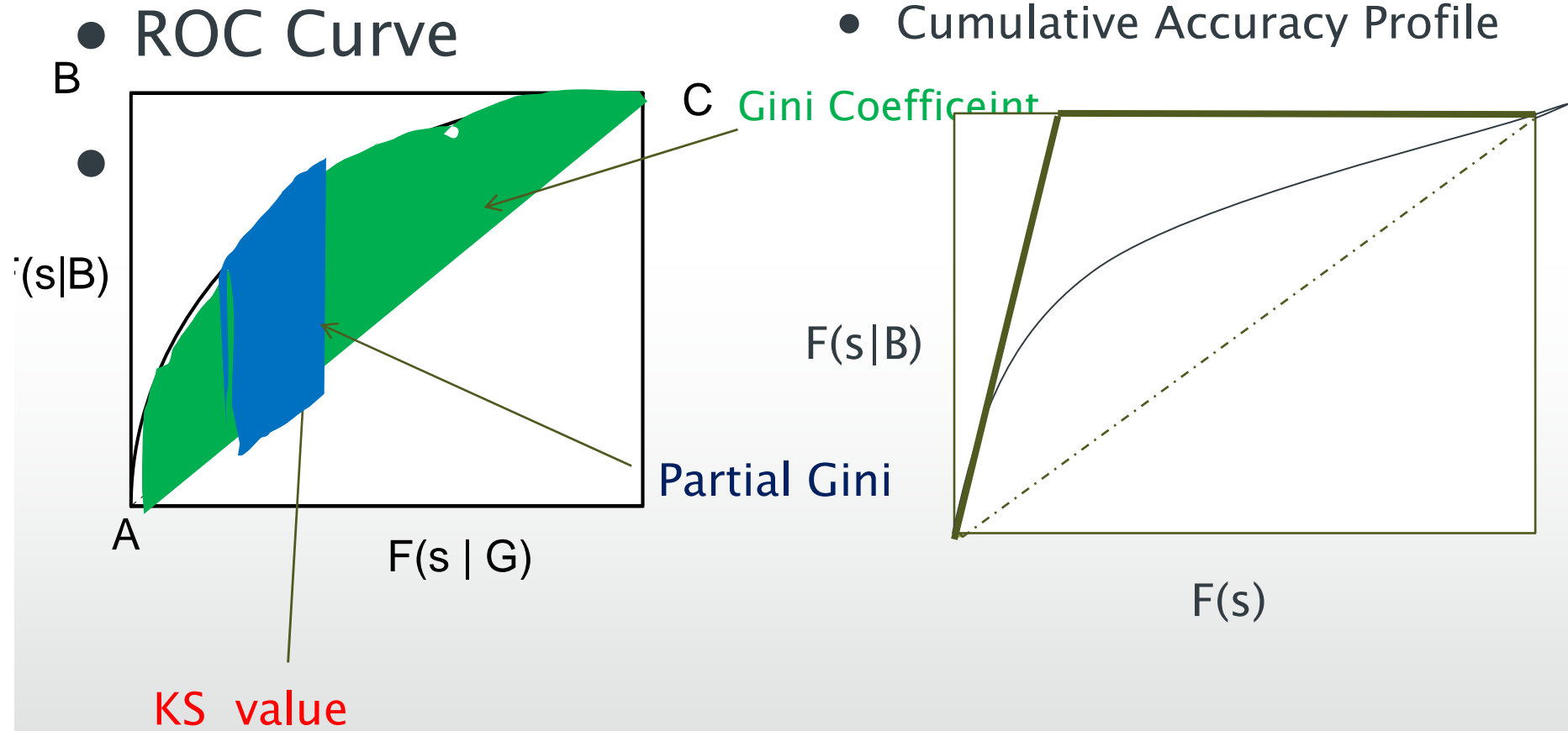
- This requires the scorecard and the cut-off score so one can implement the decisions and see how many erroneous classifications there are

Error rates
2 by 2 tables and swap sets
Hypothesis tests

Few new measures suggested

- In conferences , few papers on new ways of measuring scorecards
- Most common use is ROC curve and its measures
 - KS : max difference in % Good accepted and % Bad accepted (US)
 - Gini is the average over all cut offs in % Good accepted and %Bad accepted (RoW)
 - H measure (Hand 2007) is average over cut offs chosen with a Beta distribution
- Medical applications use partial AUROC (i.e. partial Gini)– average difference with uniform distribution over limited cut off interval
- Maybe more appropriate to use CAP curve and use partial Accuracy ratio

ROC curve and its measures



Measuring performance of LGD models

- Measure error between actual and predicted LGD values

⟨ y_i -actual LGD; \hat{y}_i -estimated LGD; \bar{y} -average LGD ⟩

– R-squared = $\frac{\text{Sum of squares regression}}{\text{Sum of squares total}} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$

– Mean Squared Error $MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}$

– Mean Absolute Deviation $MAD = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}$

- For some purposes, might want relative ranking
 - Kendall's tau – number of switches to get actual and predicted order the same
 - Spearman's rho – weight switches by number in between to get order the same

Basel wants LGD results at segment level not individual loan level

- Usually segments relate to PD values but could involve LGD values.
- So why not use basis of Hosmer –Lemeshow PD test
- Take predicted results split into N groups according to these results
 - For each group compare actual average LGD with predicted average LGD
 - Use MSE, MAD. Correlation
- Whereas individual results have $R \approx 0.3$, groups have $R \approx 0.9$

Loss Given Default and modelling the Collection Process

- Recovery Rate (RR) = Amount Recovered / Balance of loan at default
- Loss Given Default (LGD) = Amount lost / Balance of loan at default
- $RR \cong 1 - LGD$
- Although in theory $0 \leq LGD \leq 1$ can be outside this due to fees/interest.
- Implication of LGD rather than Expected Loss is that balance at default is not important
- Collection data traditionally poor but with Accord now well recorded
- Can use data to model and improve operations of collections,
 - Lots of developments in operations management ; apply those here

Current state of modelling Loss Given Default

- Corporate LGD modelling
 - Till mid 1990s used one historic value (42%)
 - Needed to price bonds so LGD value can be obtained from bond prices in market (lots of data)
 - Bonds still traded after default so can get real “LGD” from them
 - Essentially estimated by linear regression using type/seniority bonds and economic conditions
- Secured consumer lending (mortgages /car loans)
 - Loss only occurs if property repossessed
 - Estimate PPD (prob repossession given default); SAP – shortfall at repossession. Then if EAD is exposure at default
 - $EAD.LGD = PPD. SAP + (1-PPD).0$
- Unsecured consumer lending
 - Model collection process
 - Depends on actions by lender as well as uncertainty about debtors ability/willingness to pay

Modelling LGD and collections for Unsecured consumer loans

- Poor models at present. Researchers /practitioners get correlations between actual and prediction 0.2–0.5.
- Methods
 - Linear regression
 - Non linear regression : logistic, Beta, log log
 - Survival analysis: proportional hazards(COX), Accelerated Life
 - Mixed distributions
 - Logistic regression ($RR=0/RR>0$) then above
 - Neural nets
- Problems: long time period/lenders write off strategy/ impact of economic conditions
- Use dynamic programming model to describe what actions to perform and how long to perform them so as to optimise recoveries

Conclusion

- Quotation from this year's Fringe programme
- *The Edinburgh Festival and Fringe is still the first choice for performers.... and creators to come and tell their story; we are proud that it is still the place to bring your work, with opportunities to amaze, enthrall and excite audiences from both far away and close to home.*
- I hope you feel that is also the case for this conference
- Thank you and Any Questions