Society of Actuaries in Ireland

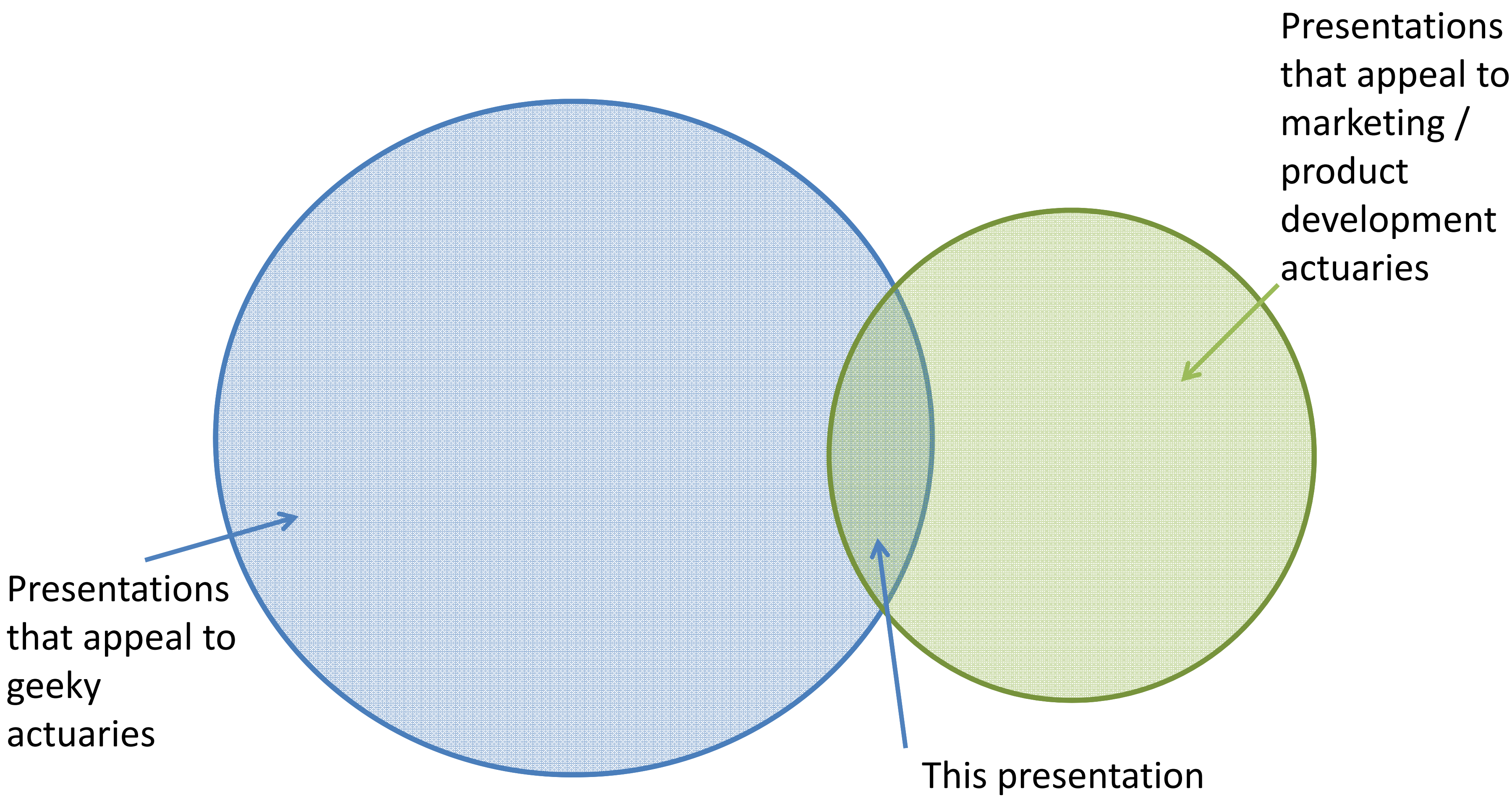
Predictive modelling around the world

28.11.13

Agenda

- Why this presentation is really interesting
- Introduction to predictive modelling
- Case studies
- Conclusions

Why this presentation is really interesting



Bulls@!t Bingo

ANALYTICS	PREDICTIVE UNDERWRITING	THE CLOUD
OPEN SOURCE TECHNOLOGY	BIG DATA	DATA MINING
MACHINE LEARNING	GLMs	CROWDSOURCING

Agenda

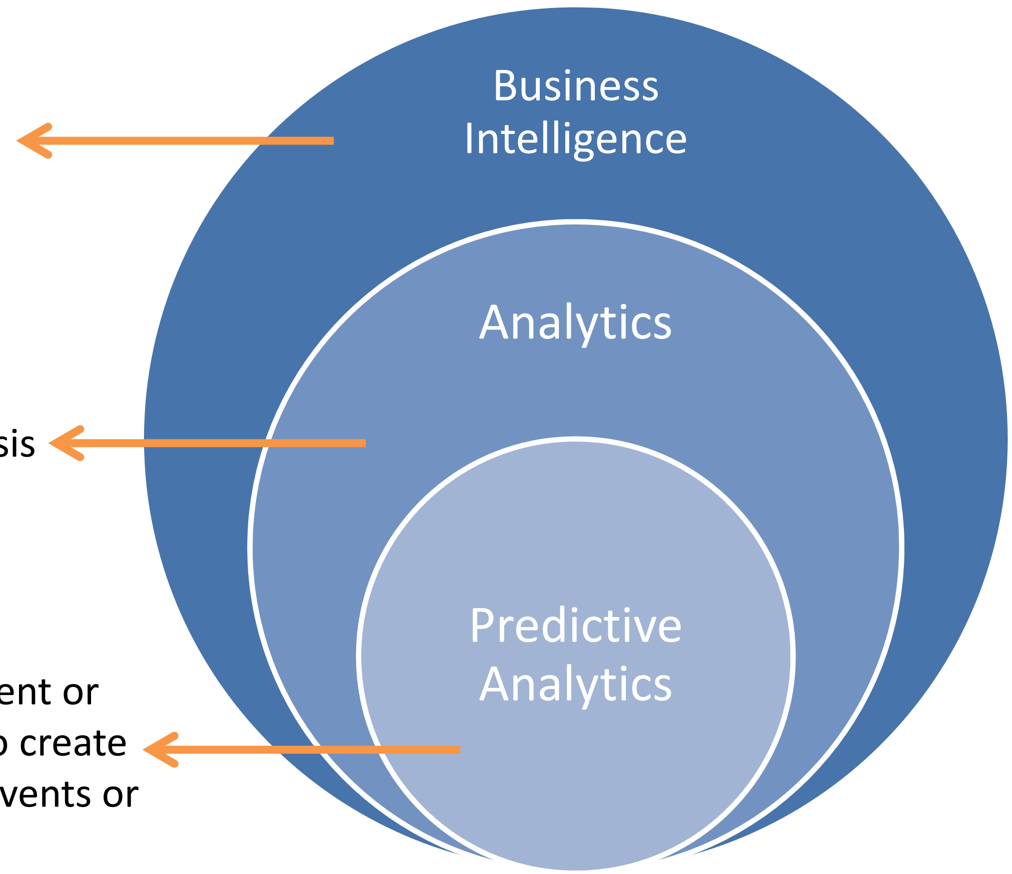
- Why this presentation is really interesting
- Introduction to predictive modelling
- Case studies
- Conclusions

What are predictive analytics?

Tools and technologies for analysing and understanding business performance

Extensive use of data with statistical and quantitative analysis

A process by which current or historical facts are used to create predictions about future events or behaviour.



What are predictive analytics?

A process, not a product

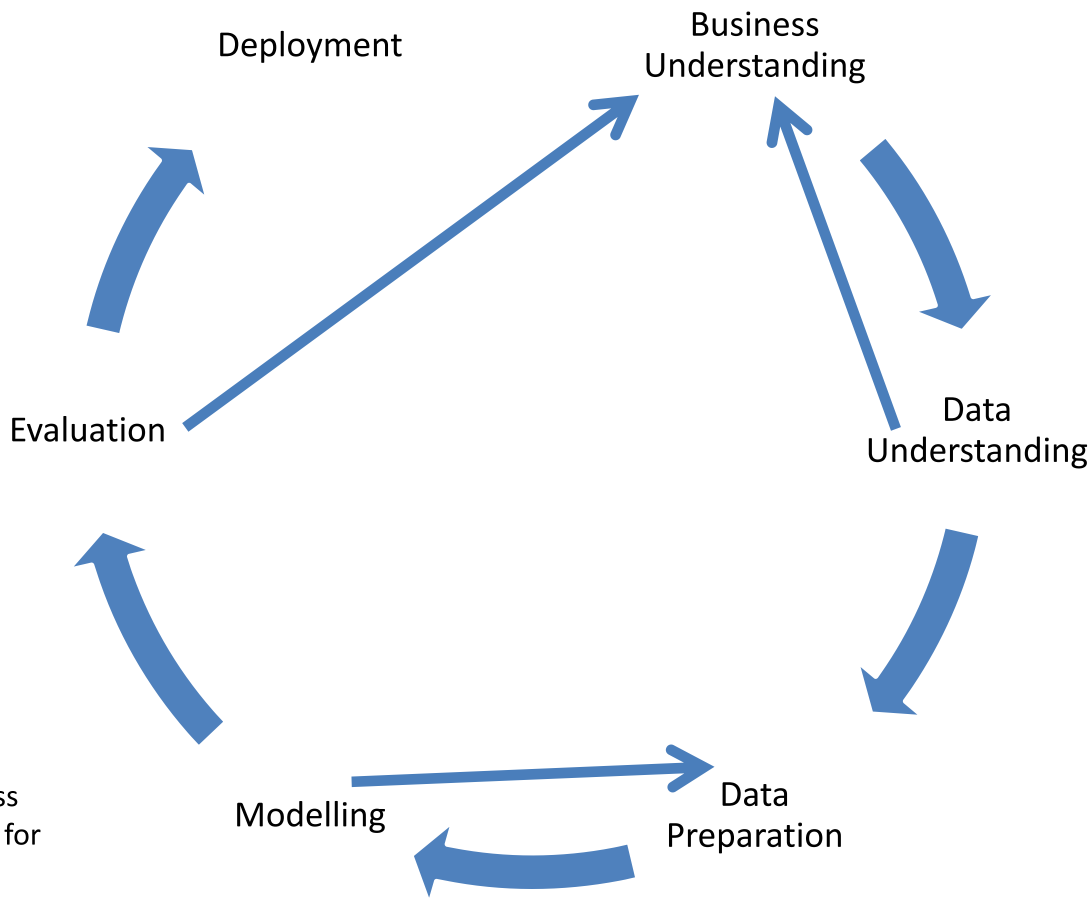
Good data is vital for success

A **process** by which current or historical **facts** are used to create **predictions** about **future events or behaviour**.

Typically predictions are created through the use of sophisticated statistical models

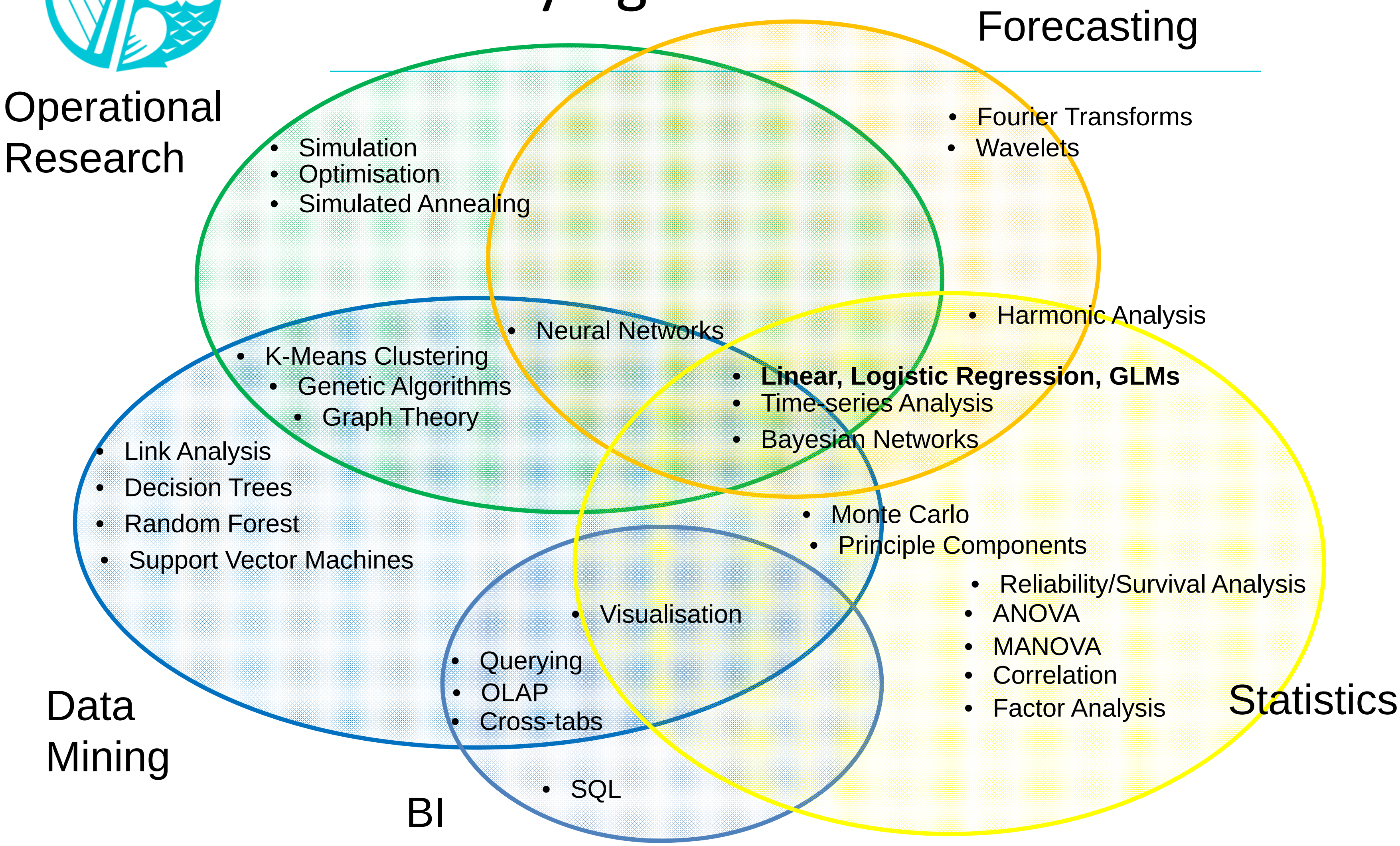
Focus on predicting probability of future events and behaviour

The process



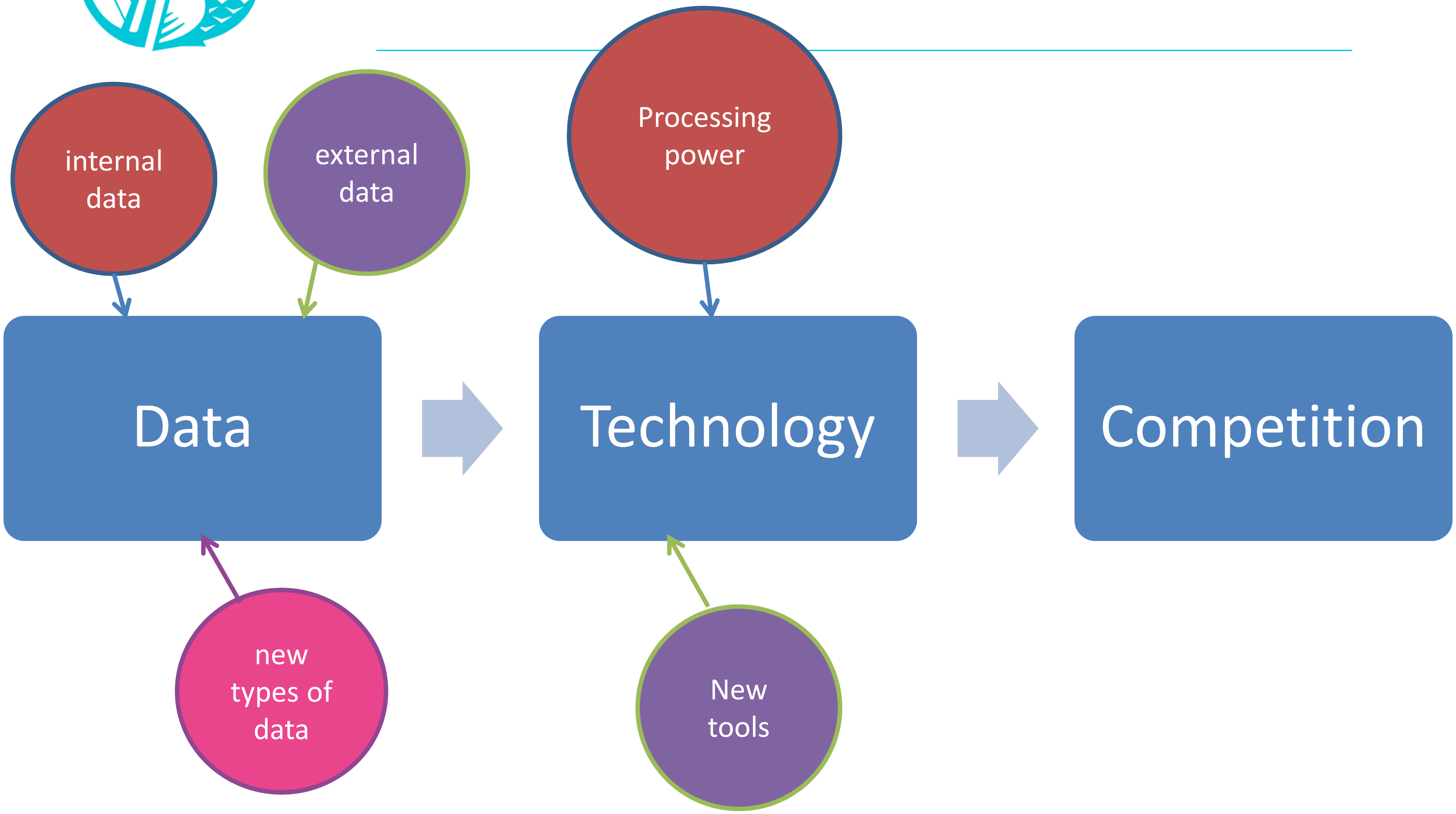
Source: adapted from Cross Industry Standard Process for Data Mining (CRISP-DM)

Underlying the models



Modified from a version presented by John Elder, www.dataminninglab.com, 2012.

Why predictive analytics?



Agenda

- Why this presentation is really interesting
- Introduction to predictive modelling
- Case studies
- Conclusions

Worldwide projects*

* Courtesy of Peter Banthorpe RGA

UK:
- Basis Setting (mortality, morbidity and lapses)
- Postcode pricing model
- Enhanced experience analysis
- Predictive underwriting on credit rating agency and bank data
- Broker Quality

Europe:
- Predictive underwriting on bancassurance data

South Africa:
- Enhanced Experience Analysis
- Predictive underwriting on bank and credit card data

India:
- Claims Fraud Prediction

Australia:
- Predictive underwriting / cross sell on bancassurance data

USA:
- Pricing override model for group LT disability
- Lapse basis
- Predictive underwriting on Non-Life data
- Term Tail Lapses
- Mortality prediction on credit rating agency data

Asia:
- Predictive underwriting on bancassurance data
- Finer price segmentation
- Propensity to buy
- Cross sell of insurance on bank data

Case Study 1: Lapse Assumptions variable annuity with GMIBs

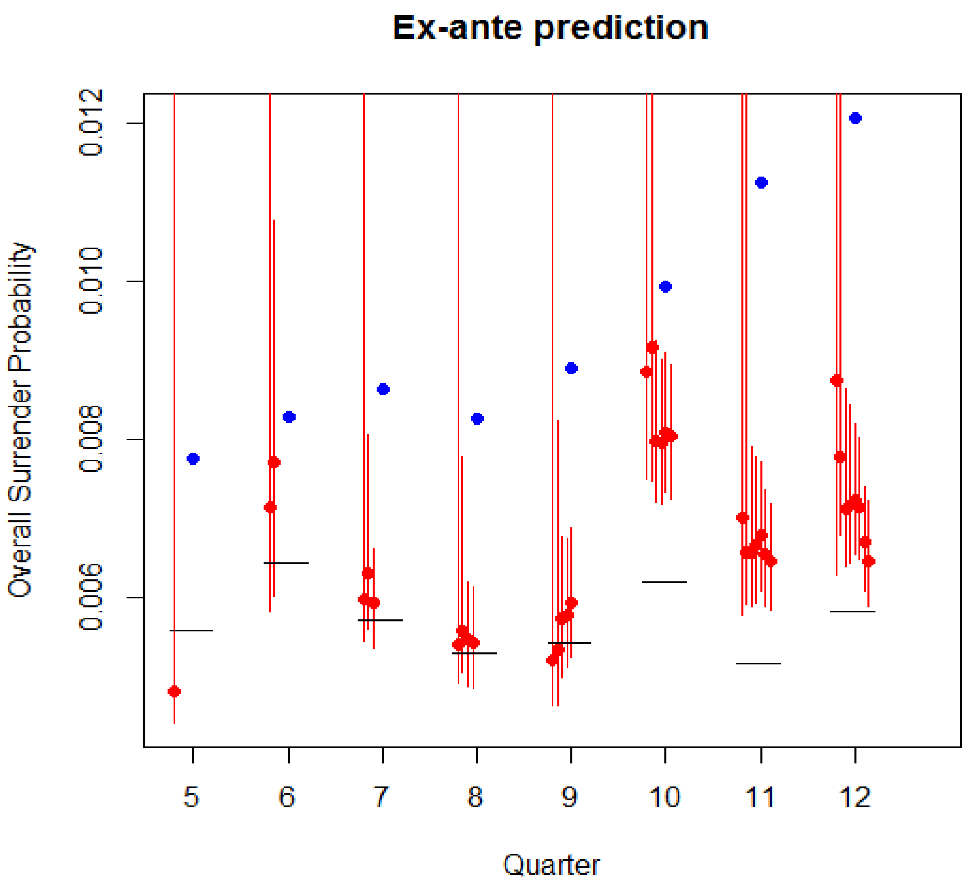
- Current lapse modeled based solely on plan type, duration, In-the-moneyness ("ITM")
- Proposal: Evaluate current model based on 12 quarters of observations
- Method: Develop an alternative statistical model based on current variables, augmented by additional policy characteristics and macroeconomic variables
- Compare predictive performance of the two models
  - Performance measured by ratio of Actual/Expected dollars lapsed in out-of-sample future period.

Predictors

Positively associated with lapse	Negatively associated with lapse
• Duration (adjusted for Surrender Charge Period)	• Annuitant age at issue
• Anniversary of issue date	• Total rider charge (bps)
• Policies sold in channel 1, 5, 6	• In-the-moneyness (ITM)
• State unemployment rate	• Held in qualified plan?
• US inflation rate	• Policy size (cumulative deposits)
• Risk aversion (low % of equity)	• Policies sold in channels 3, 8
• Past partial withdrawals	

Not significant: Gender, Surrender Charge, DB Rider, GMIB reset, past quarter S&P 500 return, 10 Yr T Note yield

Performance of predictors



- At each quarter, split data by past versus current experience
  - Actual experience (black bars)
  - Statistically predicted (red dots) expected lapse and confidence bands
  - Calculated (blue dots) expected lapse based on current assumption
- Statistical A/E outperforms current assumption in all quarters

Case Study 2: Predictive Underwriting Model

- Client: Bancassurer in Asia with large customer pool, but low penetration in life product
- Goal: to predict UW decisions on its existing customers
- Major challenges - very limited data
  - A total of about 8k-9k full UW cases
  - Target variable UW decision, with very low declined/rated cases, ~3.0%
  - Many missing values due to old time, especially for sub-STD
  - Not all information collected at the time of UW

Source: Peter Banthorpe RGA.

Modeling Approach / Key Variables

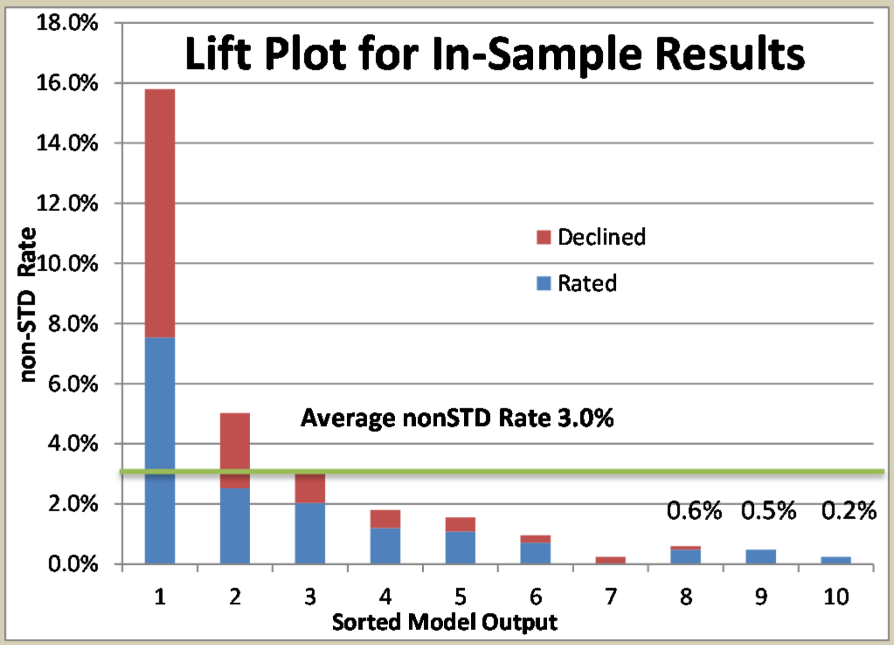
- GLM with binomial and logistic link function
- About a dozen of predictor variables that are statistically significant for prediction & readily available in client database
- Key predictor variables
  - "Positive" means the probability of being a standard rate case increases if the value goes up; otherwise, it is "Negative"

Name	Type	Note
Age_At_Entry	Numeric	Negative; less likely to qualify for STD as age goes up
Branch	Categorical	Proxy of geographic locations
Asset Under Management	Numeric	Positive; more likely to qualify for STD with large AUM
Customer_Segment	Categorical	Positive for "Gold", negative for other
Nationality	Categorical	Positive for domestic; negative for certain others

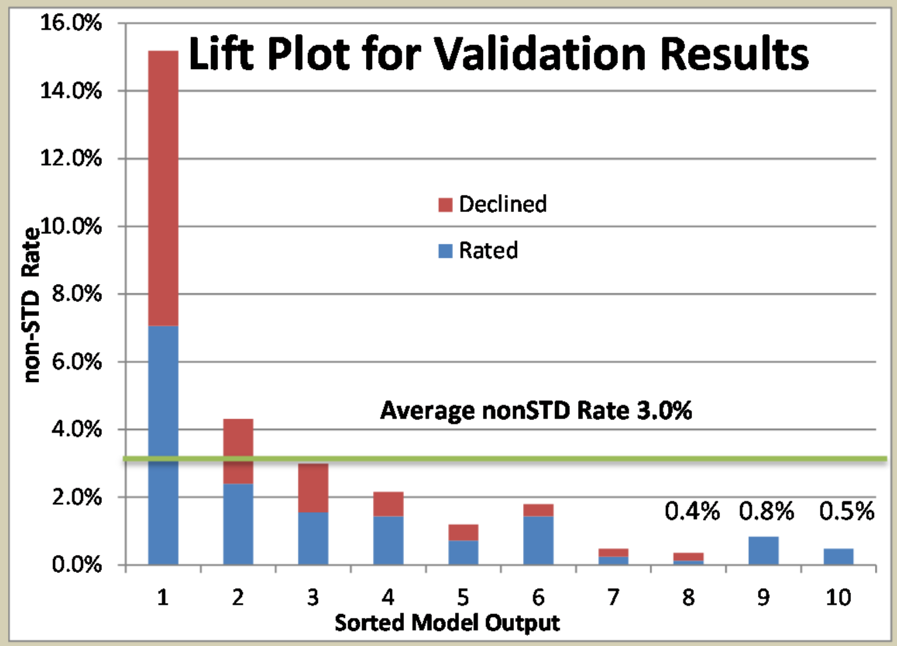
Source: Peter Banthorpe RGA.

Model Results – Lift Plots

- In-sample results show model performance under optimal condition
- May over-fit data
- 0.5% of sub-STD in top 30%



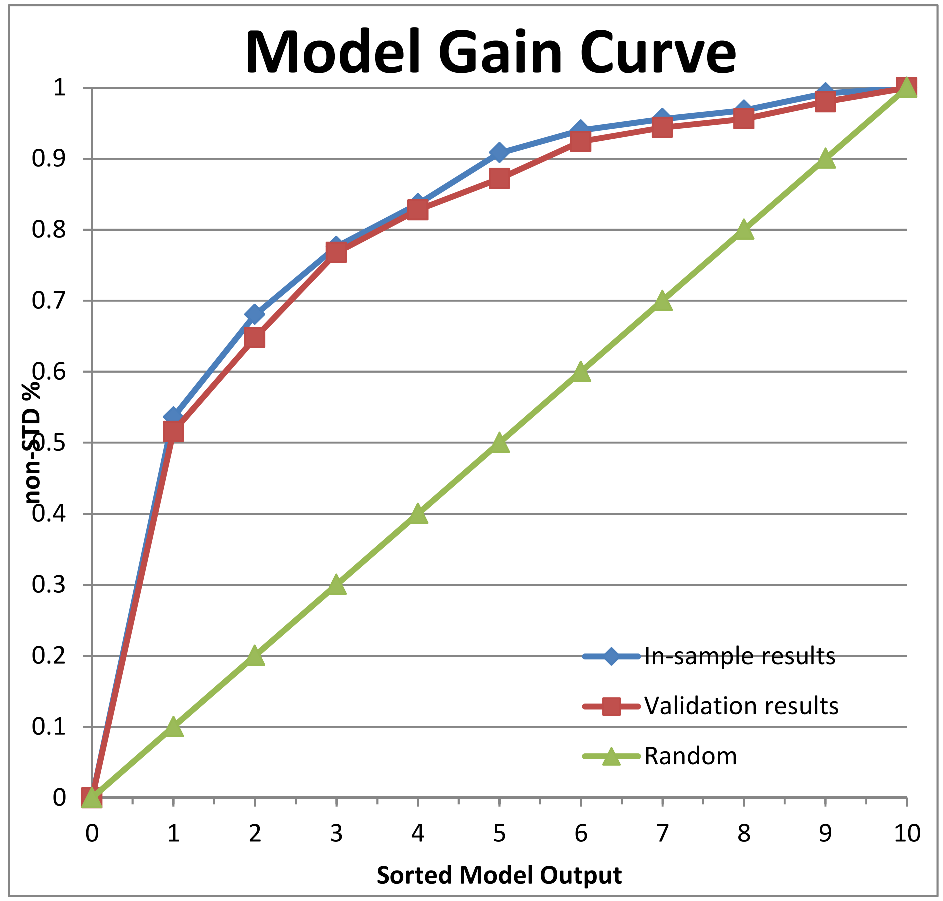
- Validation results are a better test of model performance in real business
- 0.6% sub-STD in the top 30%



Source: Peter Banthorpe RGA.

Model Results – Gain Curve

- Another way to understand model capability to differentiate STD from sub-STD
  - Best 30% of model outputs contains about 5% of total non-STD
  - Lowest 30% captures about 75% of bad risks



Source: Peter Banthorpe RGA.

Case study 3: non-life

- Predictive analytic techniques used to analyse motor insurance portfolio to:
  - Identify predictors of claims and hence model a profitability score per customer
  - Identify predictors of propensity to renew
- Allows analysis of portfolio by profitability and propensity to renew

Results

High-profit, low-retention customer segment in red

	Retention								Exposure
Loss Ratio	98.5%	97.7%	96.7%	95.8%	94.5%	91.5%	87.3%	80.8%	Total
35.0%	3.0%	1.6%	1.8%	0.9%	0.4%	0.1%	0.4%	0.5%	9%
42.7%	3.6%	2.2%	1.8%	0.8%	0.7%	0.3%	0.7%	0.6%	11%
45.8%	2.0%	3.0%	3.5%	2.4%	1.6%	1.1%	1.2%	0.5%	15%
49.5%	0.7%	2.0%	3.6%	3.4%	2.6%	1.9%	1.2%	0.6%	16%
55.5%	0.3%	1.0%	2.6%	3.4%	3.7%	2.7%	1.5%	0.7%	16%
58.2%	0.0%	0.4%	1.5%	2.7%	3.6%	4.2%	2.0%	0.8%	15%
61.4%	0.0%	0.1%	0.5%	1.4%	2.0%	2.9%	2.0%	1.0%	10%
75.4%	0.0%	0.0%	0.1%	0.7%	1.2%	2.7%	1.5%	2.0%	8%
Exposure Total	10%	10%	15%	16%	16%	16%	11%	7%	100%

Source: EagleEye Analytics

Low-profit, high-retention customer segment in red

	Retention								Exposure
Loss Ratio	98.5%	97.7%	96.7%	95.8%	94.5%	91.5%	87.3%	80.8%	Total
35.0%	3.0%	1.6%	1.8%	0.9%	0.4%	0.1%	0.4%	0.5%	9%
42.7%	3.6%	2.2%	1.8%	0.8%	0.7%	0.3%	0.7%	0.6%	11%
45.8%	2.0%	3.0%	3.5%	2.4%	1.6%	1.1%	1.2%	0.5%	15%
49.5%	0.7%	2.0%	3.6%	3.4%	2.6%	1.9%	1.2%	0.6%	16%
55.5%	0.3%	1.0%	2.6%	3.4%	3.7%	2.7%	1.5%	0.7%	16%
58.2%	0.0%	0.4%	1.5%	2.7%	3.6%	4.2%	2.0%	0.8%	15%
61.4%	0.0%	0.1%	0.5%	1.4%	2.0%	2.9%	2.0%	1.0%	10%
75.4%	0.0%	0.0%	0.1%	0.7%	1.2%	2.7%	1.5%	2.0%	8%
Exposure Total	10%	10%	15%	16%	16%	16%	11%	7%	100%

Source: EagleEye Analytics

Agenda

- Why this presentation is really interesting
- Introduction to predictive modelling
- Case studies
- Conclusions

Key messages

- Data driven process
- Broad potential applications for insurance
- Non-Life is way ahead of us
- Simplified underwriting a key area of focus (but there are many more applications)
- Not an off-the shelf solution
  - Customised, based on specific data and specific needs
  - No two exercises are the same – flexibility of approach is key