Session 044: Incorporating Predictive Analytics in the Insurance Value Chain

SOCIETY OF ACTUARIES
Antitrust Compliance Guidelines

Active participation in the Society of Actuaries is an important aspect of membership. While the positive contributions of professional societies and associations are well-recognized and encouraged, association activities are vulnerable to close antitrust scrutiny. By their very nature, associations bring together industry competitors and other market participants.

The United States antitrust laws aim to protect consumers by preserving the free economy and prohibiting anti-competitive business practices; they promote competition. There are both state and federal antitrust laws, although state antitrust laws closely follow federal law. The Sherman Act, is the primary U.S. antitrust law pertaining to association activities. The Sherman Act prohibits every contract, combination or conspiracy that places an unreasonable restraint on trade. There are, however, some activities that are illegal under all circumstances, such as price fixing, market allocation and collusive bidding.

There is no safe harbor under the antitrust law for professional association activities. Therefore, association meeting participants should refrain from discussing any activity that could potentially be construed as having an anti-competitive effect. Discussions relating to product or service pricing, market allocations, membership restrictions, product standardization or other conditions on trade could arguably be perceived as a restraint on trade and may expose the SOA and its members to antitrust enforcement procedures.

While participating in all SOA in person meetings, webinars, teleconferences or side discussions, you should avoid discussing competitively sensitive information with competitors and follow these guidelines:

- **Do not** discuss prices for services or products or anything else that might affect prices
- **Do not** discuss what you or other entities plan to do in a particular geographic or product markets or with particular customers.
- **Do not** speak on behalf of the SOA or any of its committees unless specifically authorized to do so.
- **Do** leave a meeting where any anticompetitive pricing or market allocation discussion occurs.
- **Do** alert SOA staff and/or legal counsel to any concerning discussions
- **Do** consult with legal counsel before raising any matter or making a statement that may involve competitively sensitive information.

Adherence to these guidelines involves not only avoidance of antitrust violations, but avoidance of behavior which might be so construed. These guidelines only provide an overview of prohibited activities. SOA legal counsel reviews meeting agenda and materials as deemed appropriate and any discussion that departs from the formal agenda should be scrutinized carefully. Antitrust compliance is everyone's responsibility; however, please seek legal counsel if you have any questions or concerns.

Presentation Disclaimer

Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the participants individually and, unless expressly stated to the contrary, are not the opinion or position of the Society of Actuaries, its cosponsors or its committees. The Society of Actuaries does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented. Attendees should note that the sessions are audio-recorded and may be published in various media, including print, audio and video formats without further notice.

SOCIETY OF ACTUARIES

Session Presented By:

Predictive Analytics and Futurism Section

Provides opportunities for actuaries to deepen their understanding of predictive analytics and emerging technologies relevant to the future of the actuarial profession and insurance industry.

Section Developed Content & Benefits

Predictive Analytics and Futurism Newsletter

Discusses futurism and the latest predictive analytics trends. Published three times a year. Digital editions now available.

SOA Meetings and Seminars

Section developed content presented during meeting sessions and seminars.

Podcasts

Expert led technical podcasts exploring the latest predictive analytics concepts and techniques.

Webcasts

Discounts on section developed webcasts. Free access to section created webcasts over one-year old.

Join the PAF Section Today! ***SOA.org/PAF***

CONNECT WITH SECTION MEMBERS

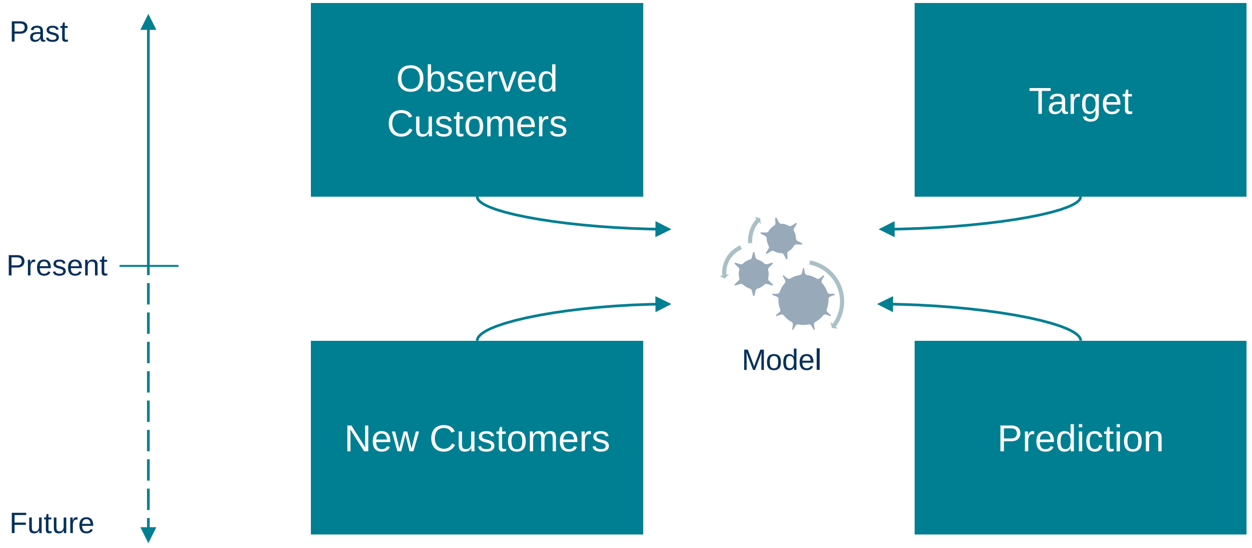
Incorporating Predictive Analytics in the Insurance Value Chain

Adnan Haque

October 28, 2019

What is predictive analytics?

Capture relationships between explanatory variables and historical outcomes



Types of predictive analytics

The vast majority of predictive analytics in insurance is still at the first rung

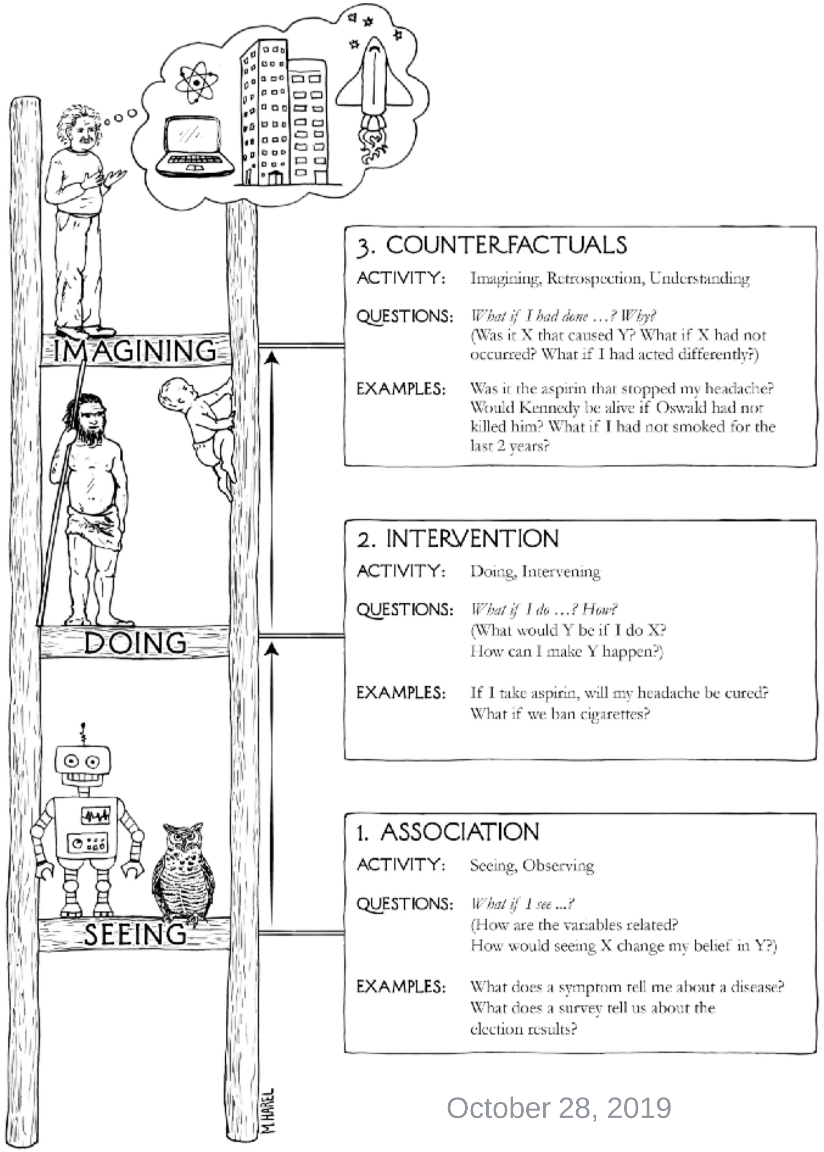


Image from The Book of Why: The New Science of Cause and Effect by Judea Pearl

DATA & AI LANDSCAPE 2019

INFRASTRUCTURE

- HADOOP ON-PREMISE
- HADOOP IN THE CLOUD
- STREAMING / IN-MEMORY
- NoSQL DATABASES
- NewSQL DATABASES
- GRAPH DBs
- MPP DBs
- CLOUD EDW
- SERVERLESS
- DATA TRANSFORMATION
- DATA INTEGRATION
- DATA GOVERNANCE
- MGMT / MONITORING
- STORAGE
- CLUSTER SVCS
- DATA GENERATION & LABELLING
- AI OPS
- GPU DBs & CLOUD
- HARDWARE

ANALYTICS & MACHINE INTELLIGENCE

- DATA ANALYST PLATFORMS
- DATA SCIENCE PLATFORMS
- BI PLATFORMS
- VISUALIZATION
- MACHINE LEARNING
- COMPUTER VISION
- HORIZONTAL AI
- SPEECH & NLP
- SEARCH
- LOG ANALYTICS
- SOCIAL ANALYTICS
- WEB / MOBILE / COMMERCE ANALYTICS

APPLICATIONS – ENTERPRISE

- SALES
- MARKETING - B2B
- MARKETING - B2C
- CUSTOMER EXPERIENCE / SERVICE
- ENTERPRISE PRODUCTIVITY
- HUMAN CAPITAL
- LEGAL
- REGTECH & COMPLIANCE
- PARTNERSHIPS
- FINANCE
- BACK OFFICE AUTOMATION & RPA
- SECURITY

APPLICATIONS – INDUSTRY

- ADVERTISING
- EDUCATION
- REAL ESTATE
- GOV'T
- INTELLIGENCE
- FINANCE - INVESTING
- FINANCE - LENDING
- INSURANCE
- HEALTHCARE
- LIFE SCIENCES
- TRANSPORTATION
- AGRICULTURE
- COMMERCE
- INDUSTRIAL
- OTHER

CROSS-INFRASTRUCTURE/ANALYTICS

OPEN SOURCE

- FRAMEWORKS
- QUERY / DATA FLOW
- DATA ACCESS & DATABASES
- ORCHESTRATION & MGMT
- STREAMING & MESSAGING
- STAT TOOLS & LANGUAGES
- AI OPS & INFRA
- AI / MACHINE LEARNING / DEEP LEARNING
- SEARCH
- LOGGING & MONITORING
- VISUALIZATION
- COLLABORATION
- SECURITY

DATA SOURCES & APIs

- HEALTH
- IOT
- FINANCIAL & ECONOMIC DATA
- AIR / SPACE / SEA
- PEOPLE / ENTITIES
- LOCATION INTELLIGENCE
- OTHER

DATA RESOURCES

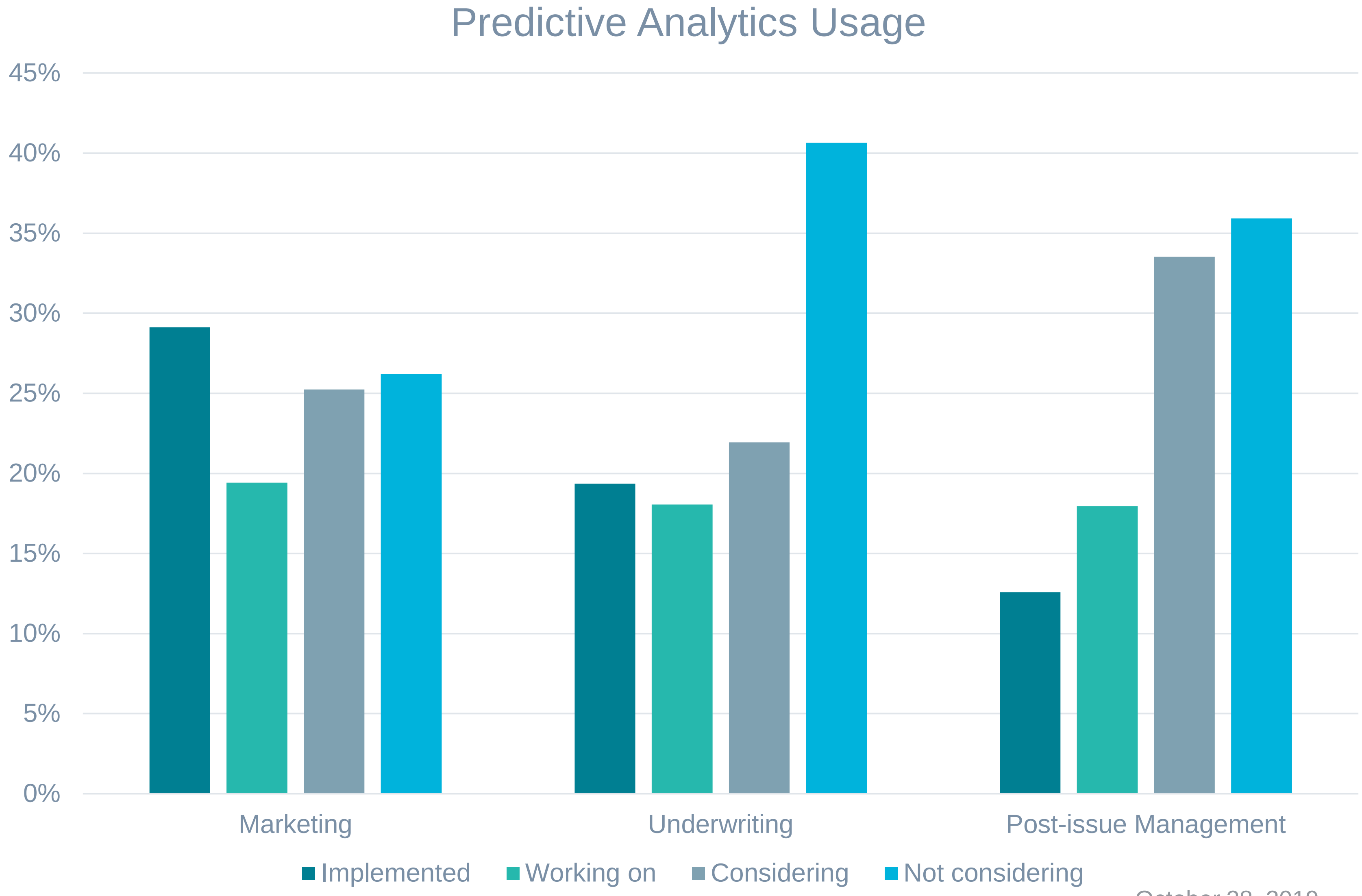
- DATA SERVICES
- INCUBATORS & SCHOOLS
- RESEARCH

July 16, 2019 - FINAL 2019 VERSION

mattturck.com/data2019

FIRSTMARK
EARLY STAGE VENTURE CAPITAL

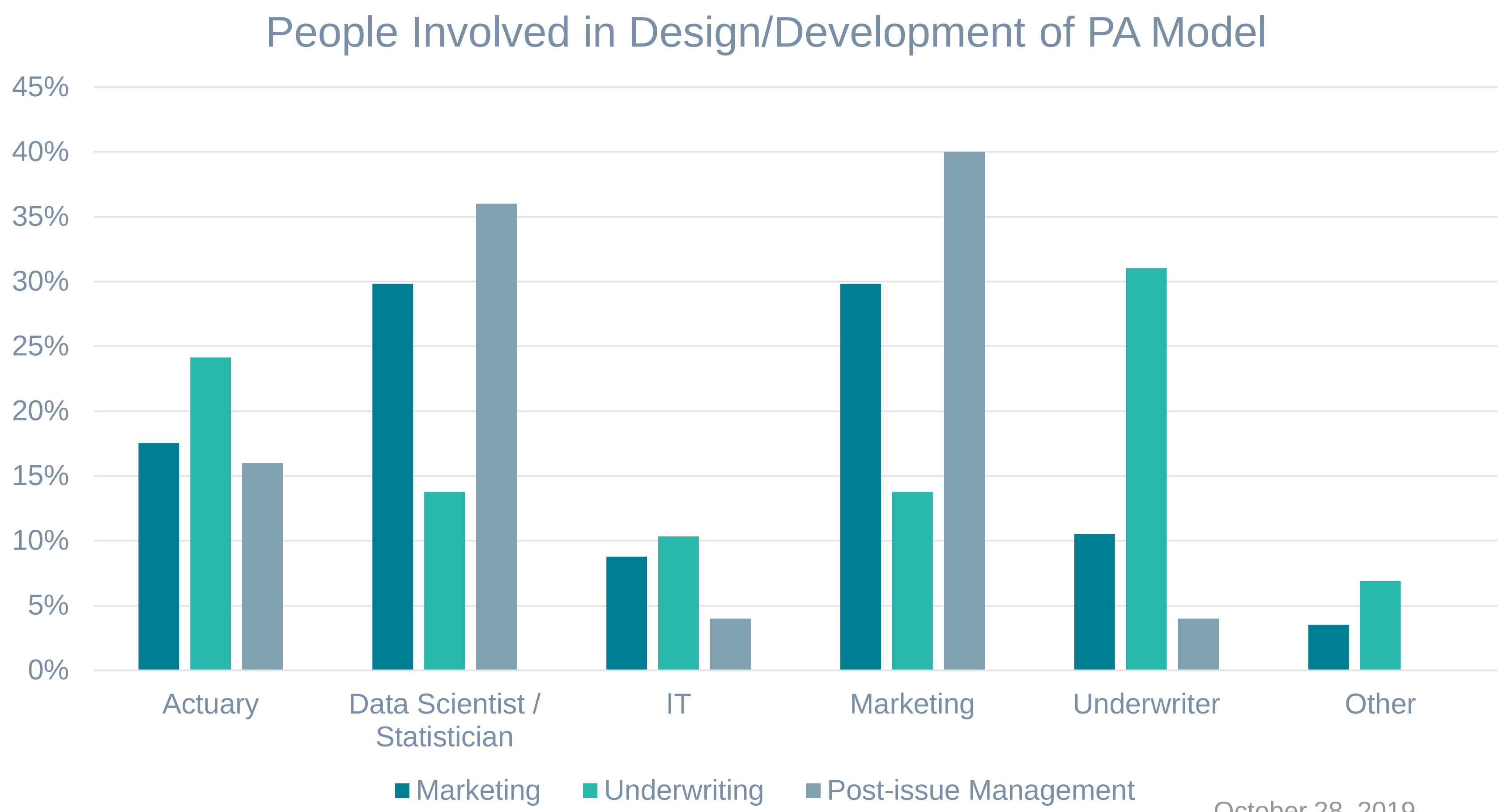
Predictive analytics usage in life insurance



Predictive Analytics and Accelerated Underwriting Survey Report May 2017

Areas involved in predictive analytics development

25% of predictive analytics projects are developed with external partners



Predictive Analytics and Accelerated Underwriting Survey Report May 2017

Predictive analytics use cases

Accelerate underwriting

- Eliminate evidence
- Automate decisions

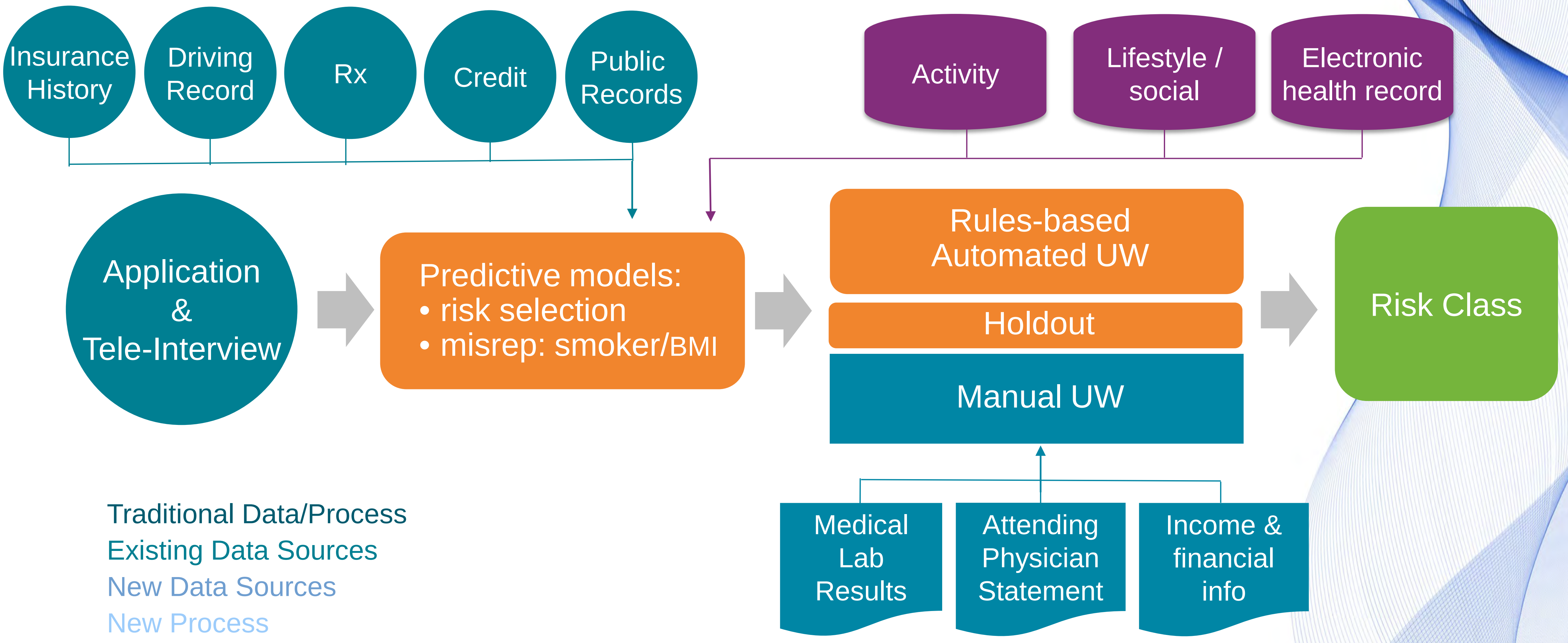
Price more accurately

- Incorporate more factors into mortality / morbidity prediction
- Provide finer segmentation or even individual pricing

Drive sales and marketing with data

- New models of IT
- Target for both marketing and risk

Accelerated underwriting landscape



Smoker detection

Objective: Manage incremental mortality risk as misrepresentation of smoking behavior increases in non-fluid program

Key predictors:

- Application: age, gender, height, weight, term & face amount, illegal drug/alcohol
- Census: occupations, population density, etc.

Two-Model Approach:

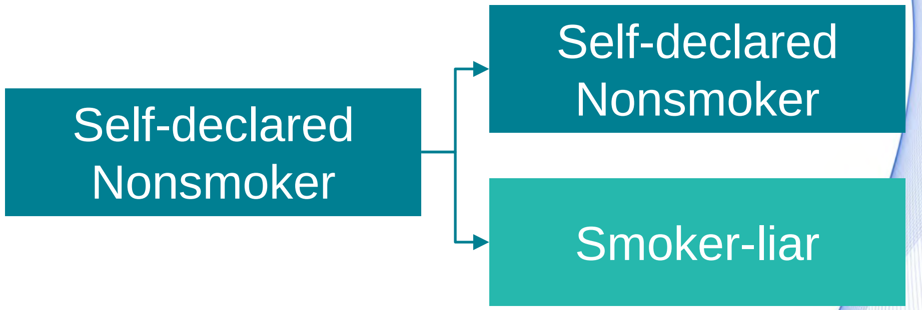
Model Smoker-Liar (SL):

- Based on all information in the app, incl. self-reported smoking history and medical drill down questions

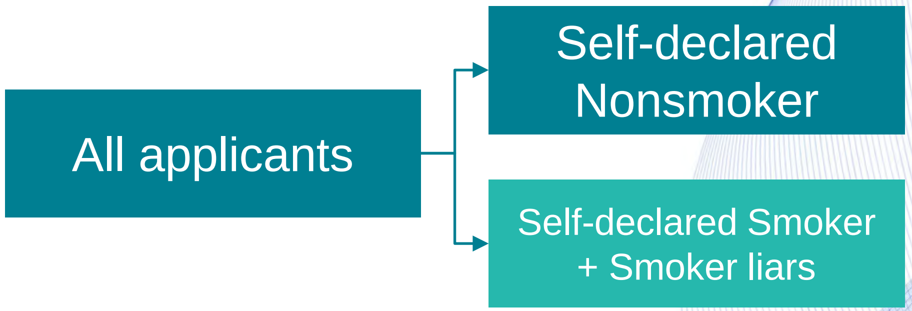
Model Smoker-All (SA):

- Ignores all self-reported smoking questions
- Heavily relies on public data sources

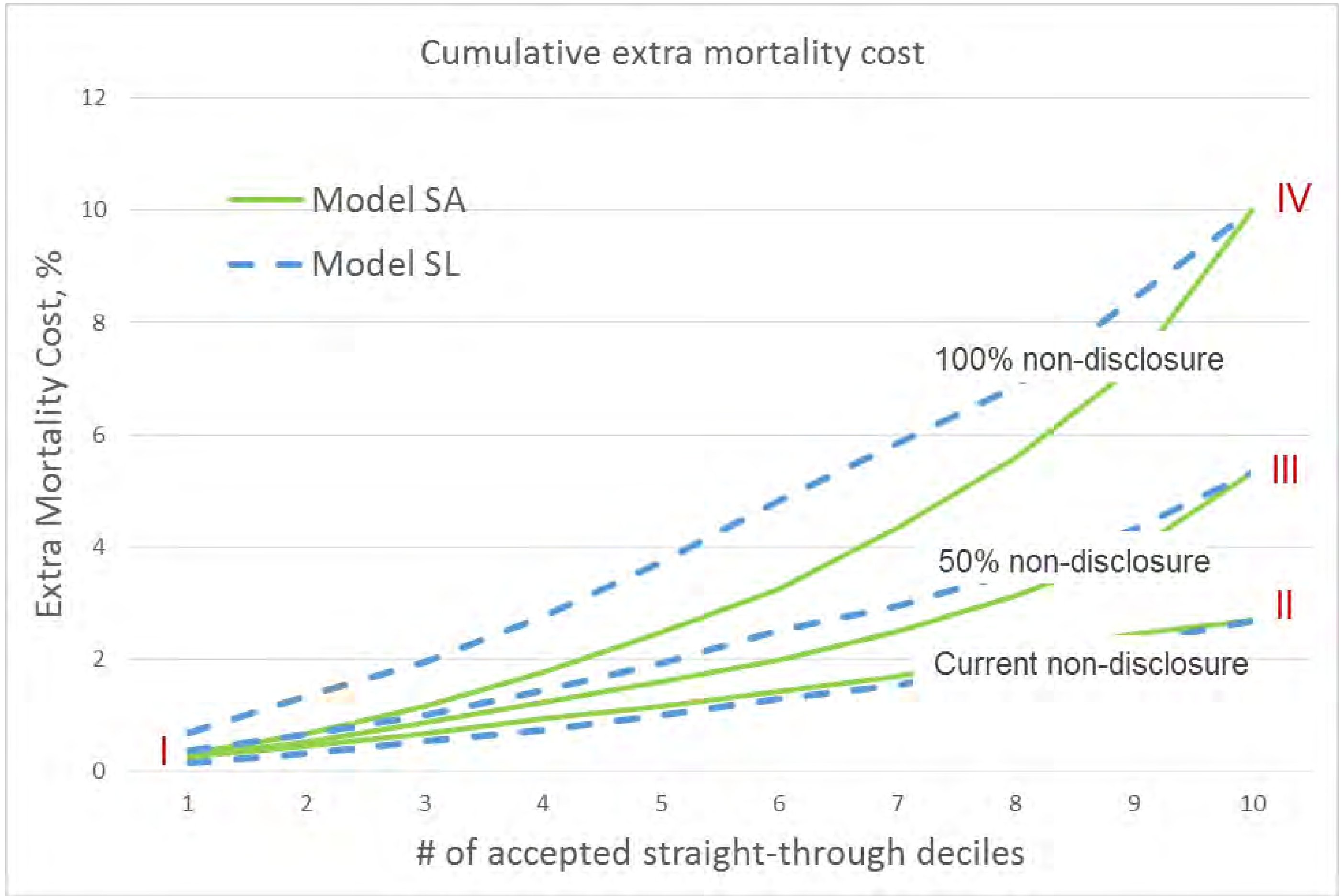
Smoker-liar model:



Smoker-all model:



Smoker detection



	Fluid Test	Non-disclosure	Extra Mortality*
I	All	25%	0%
II	None	25%	2.7%
III	None	50%	5.3%
IV	None	100%	10%

- Currently all applicants are sent for fluid tests; extra mortality is 0%
- When fluids are eliminated without routing likely smokers for tests, mortality will increase
- At current self-disclosure, Model SL minimizes extra mortality cost (slightly)
- As non-disclosure increases, Model SA minimizes extra mortality cost

**Calculated as smoker liar rate * 100% (mortality multiplier for smokers)*
Extra mortality figures are illustrative

Munich RE

Underwriting class prediction

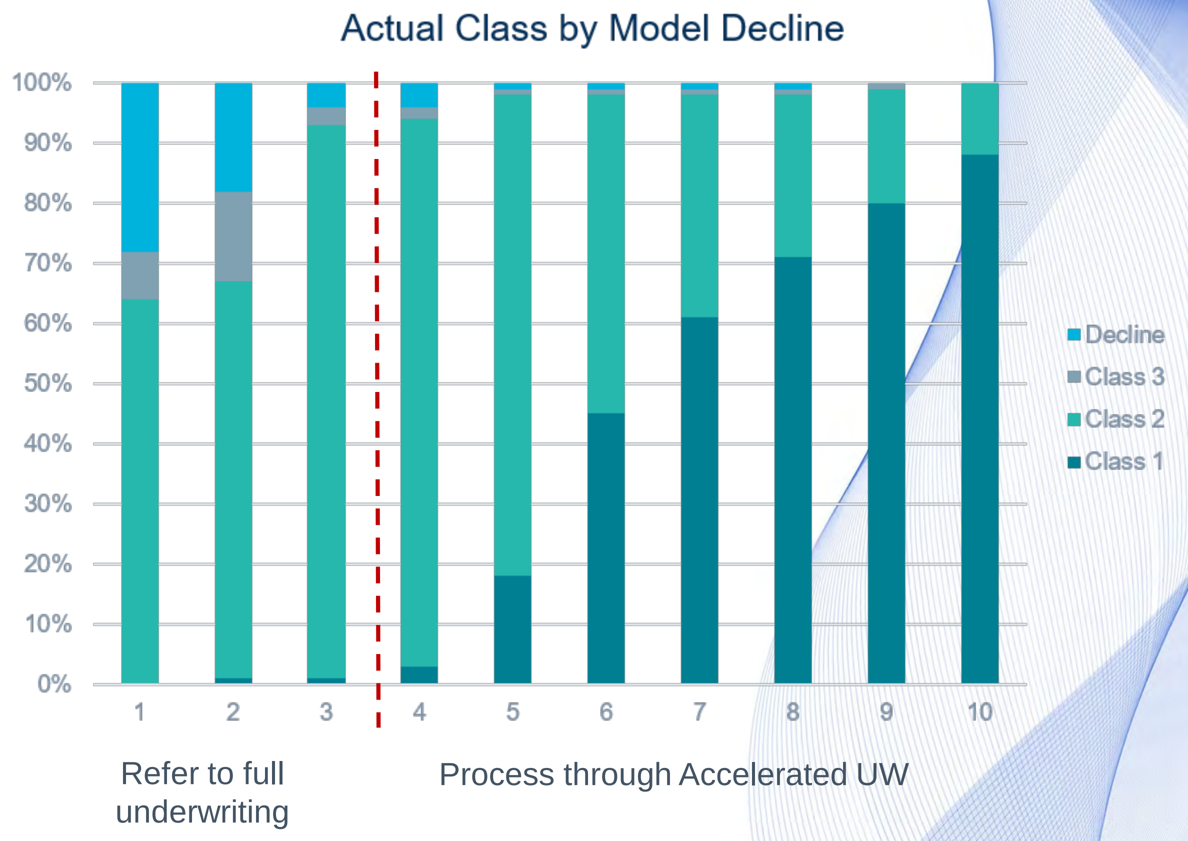
Objective: Manage incremental mortality risk in a non-fluid program by using a predictive model to identify the best fitting class for the case

Model Predicted Probabilities

Case	Actual UW Class	Class 1	Class 2	Class 3	Decline	Predicted Class	Action
1	Class 1	96%	3%	1%	0%	Class 1	AUW
2	Declined	2%	29%	47%	23%	Class 3	FUW
3	Class 2	11%	68%	16%	5%	Class 2	AUW
4	Class 1	73%	16%	8%	3%	Class 1	AUW
5	Class 3	63%	6%	29%	2%	Class 1	AUW

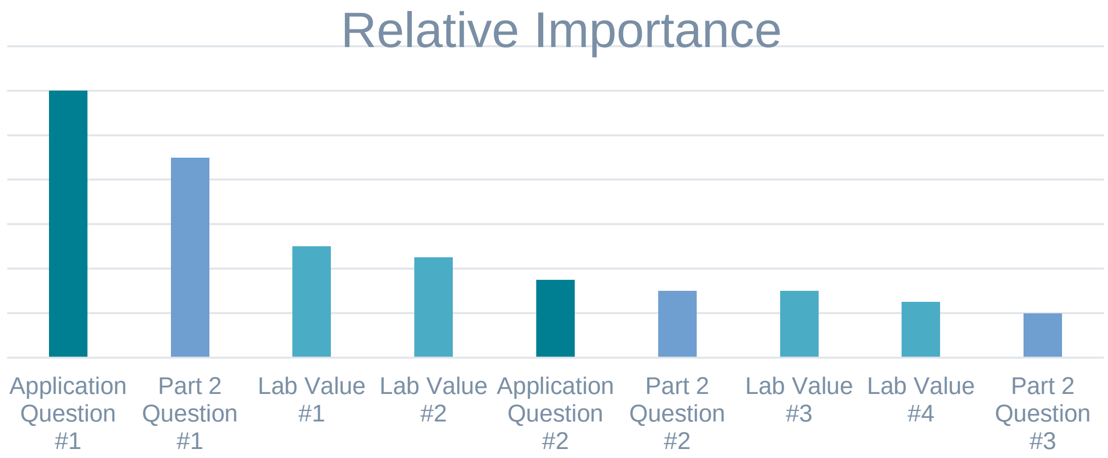
Set thresholds to triage cases

- Maximum score for worse cases to be sent to full underwriting
- Minimum score for best cases to be allowed into accelerated underwriting

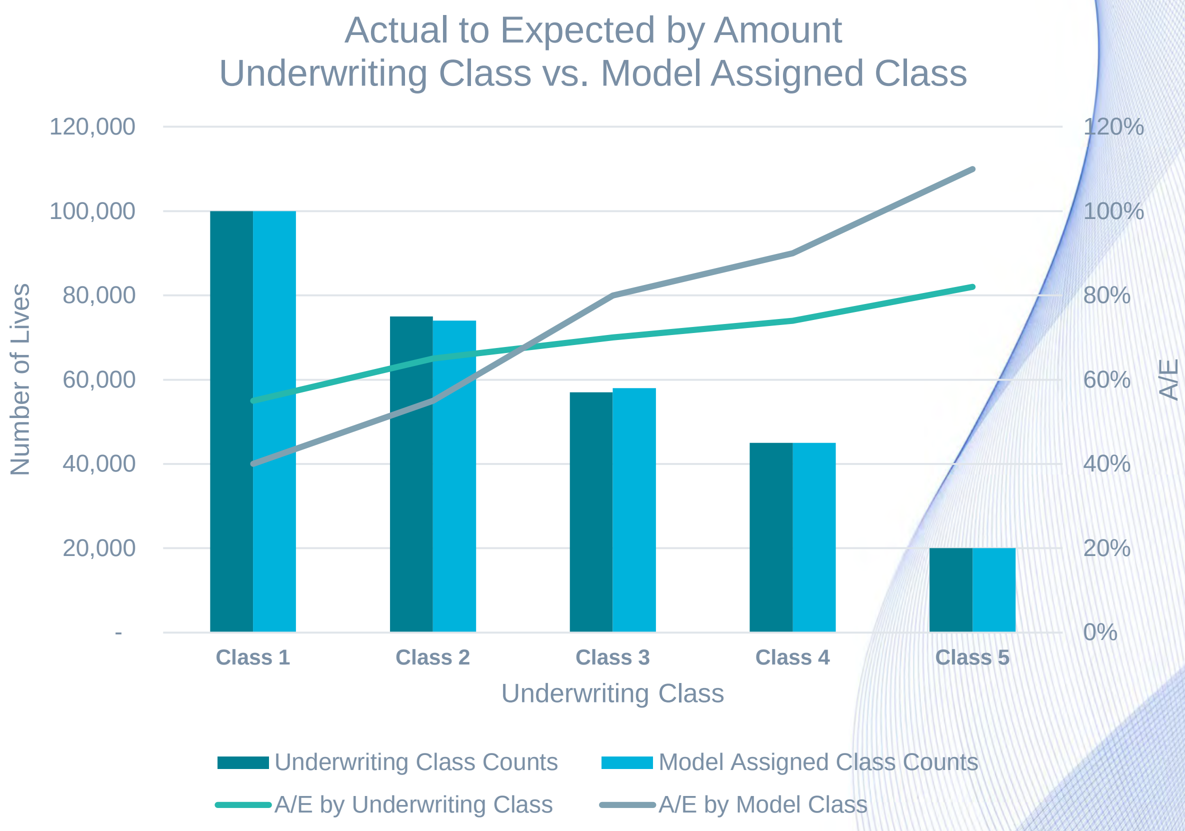


Mortality prediction

Objective: Discover misalignment between underwriting risk assessment and realized mortality by modeling claims directly



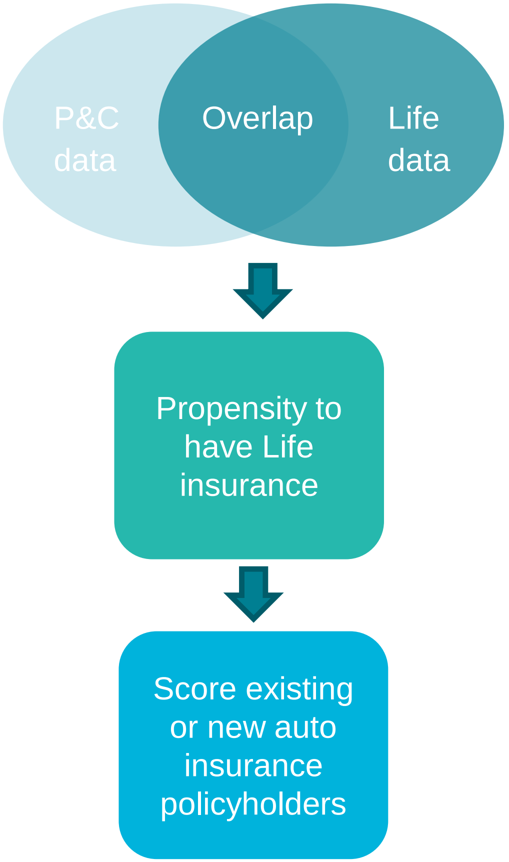
- Assign risk classes using the mortality model
- Model class assignment is controlled for age, gender, and smoking status
- Model classes have same or greater A/E differentiation than underwriting class

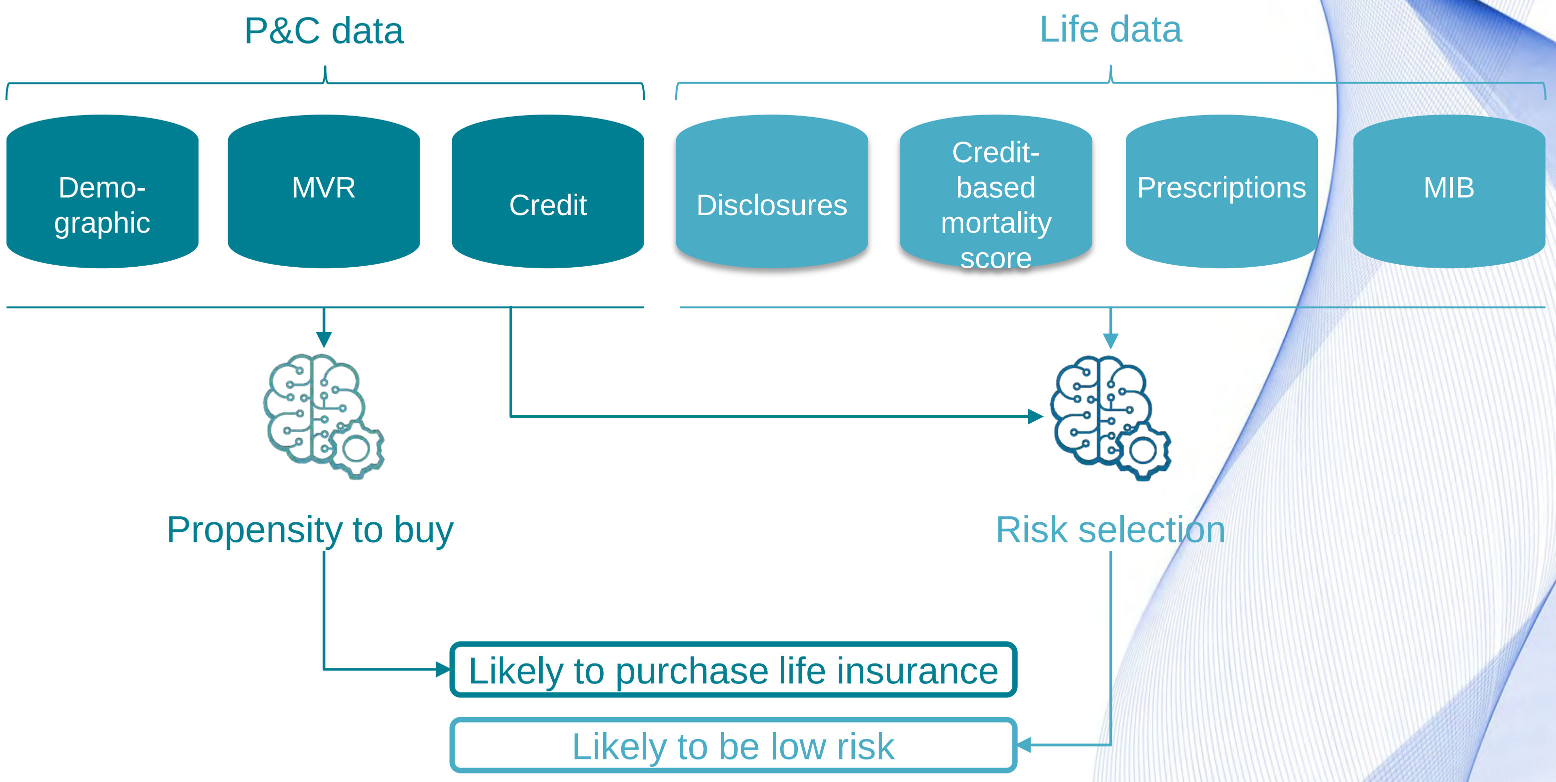


Cross sell

Objective: Cross sell life insurance to existing P&C customers

1. Assemble Data
2. Build Model
3. Score prospects





Munich RE

Proactive lapse management

Objective: Reduce lapses by identifying customers likely to lapse in the next 90 days.

Policy

Agent

Financial transactions

Service

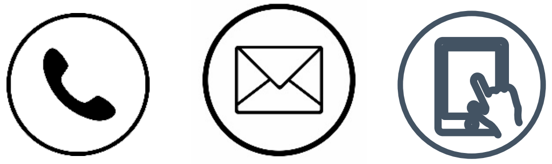
List of policyholders most likely to surrender

- Adam Smith - $10k
- Bob Joseph - $250k
- Greg Michaels - $100k
- Adnan Haque - $1mm

■ Treatment

■ No treatment

Tailored outreach



Key predictors:

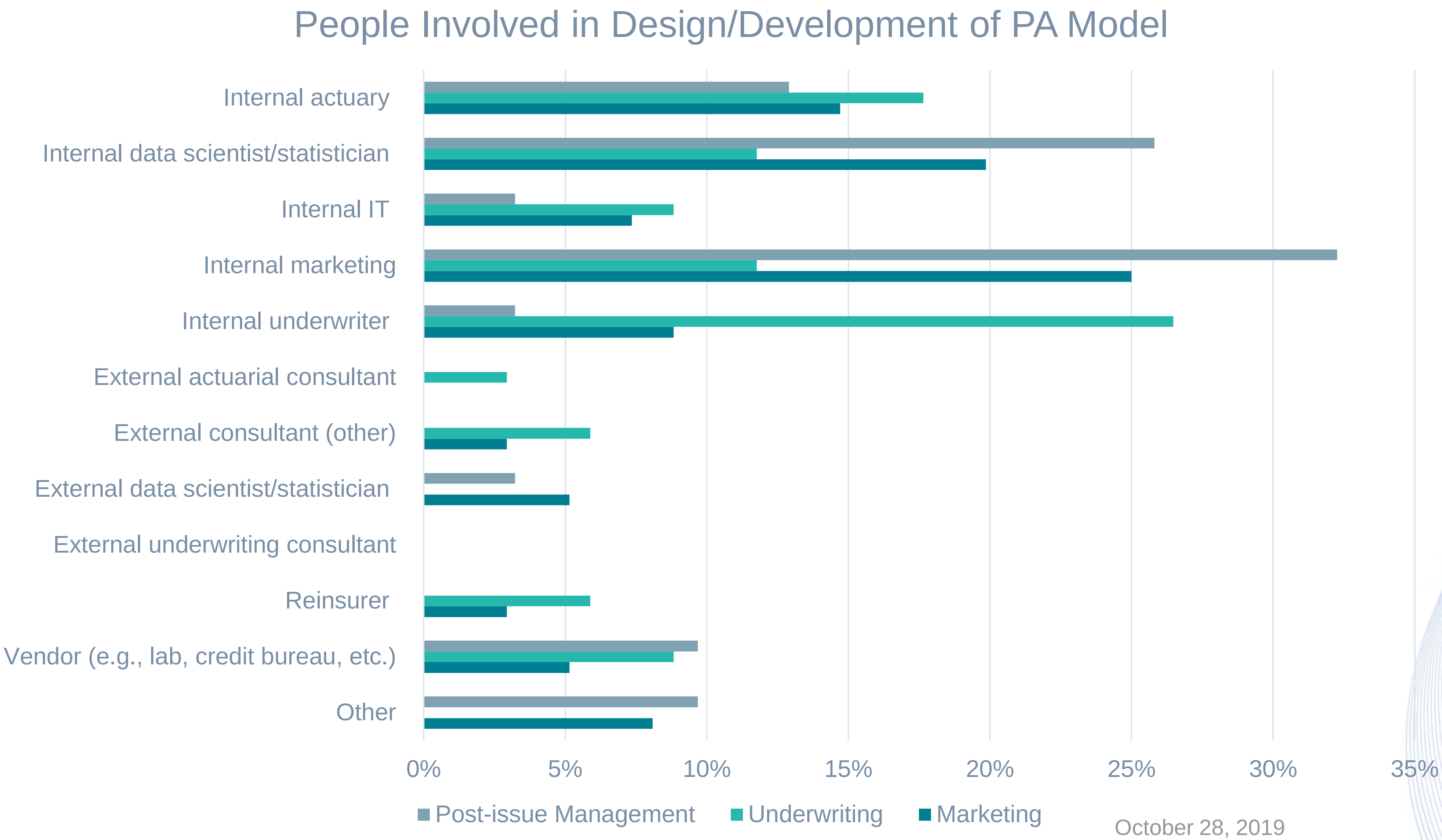
- Static: demographic, policy features, etc.
- Dynamic: disbursements, account transfers, service interactions, agent production level, etc.

Approach:

- Pro-active outreach to in-force customers that are likely to surrender
- Monthly retraining of model to identify new patterns associated with lapse

Appendix

Predictive analytics - who



Predictive Analytics and Accelerated Underwriting Survey Report May 2017

Thank you!

2019 ANNUAL MEETING & EXHIBIT

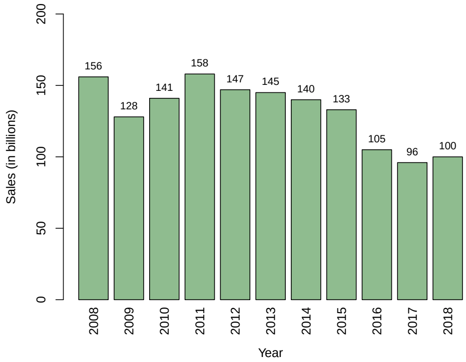
Metamodels and the valuation of large variable annuity portfolios

Emiliano A. Valdez, PhD, FSA
joint work with Guojun Gan, PhD, FSA

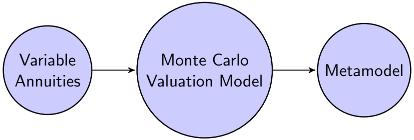
University of Connecticut

SOCIETY OF ACTUARIES®

Efficient valuation of large variable annuity portfolios



1. A challenge

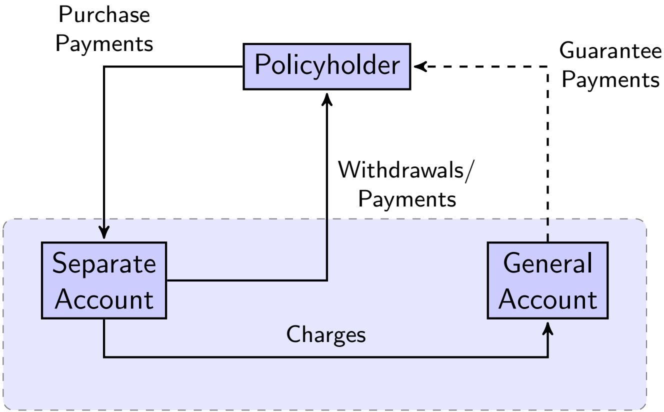


2. A metamodeling approach

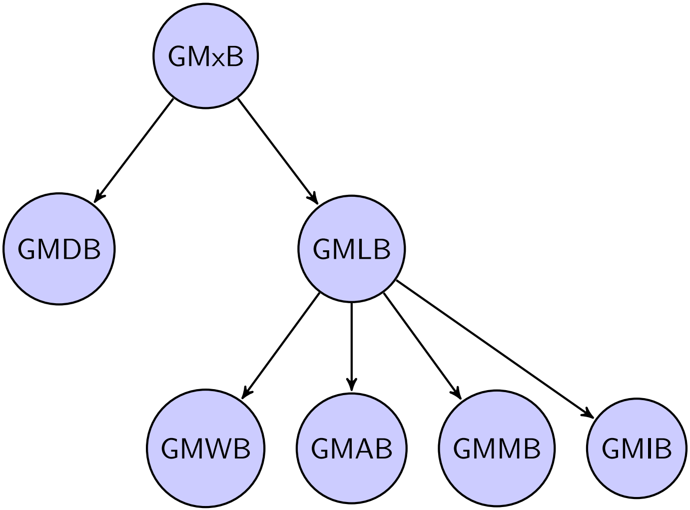
3. Numerical results

What is a variable annuity?

A variable annuity is a retirement product, offered by an insurance company, that gives you the option to select from a variety of investment funds and then pays you retirement income, the amount of which will depend on the investment performance of funds you choose.



Variable annuities come with guarantees



Insurance companies have to make guarantee payments under bad market conditions

Example (An immediate variable annuity with GMWB)

- Total investment and initial benefits base: $100,000
- Maximum annual withdrawal: $8,000

Policy Year	INV Return	Fund Before WD	Annual WD	Fund After WD	Remaining Benefit	Guarantee CF
1	-10%	90,000	8,000	82,000	92,000	0
2	10%	90,200	8,000	82,200	84,000	0
3	-30%	57,540	8,000	49,540	76,000	0
4	-30%	34,678	8,000	26,678	68,000	0
5	-10%	24,010	8,000	16,010	60,000	0
6	-10%	14,409	8,000	6,409	52,000	0
7	10%	7,050	8,000	0	44,000	950
8	r	0	8,000	0	36,000	8,000
⋮	⋮	⋮	⋮	⋮	⋮	⋮
12	r	0	8,000	0	4,000	8,000
13	r	0	4,000	0	0	4,000

Dynamic hedging

Dynamic hedging is a popular approach to mitigate the financial risk, but

- Dynamic hedging requires calculating the dollar Deltas of a portfolio of variable annuity policies within a short time interval.
- The value of the guarantees cannot be determined by closed-form formula.
- The Monte Carlo simulation model is time-consuming.

There is also the additional computational issue related to reflect the effect of dynamic hedging in (quarterly) financial reporting.

Use of Monte Carlo method

Using the Monte Carlo method to value large variable annuity portfolios is time-consuming:

Example (Valuing a portfolio of 100,000 policies)

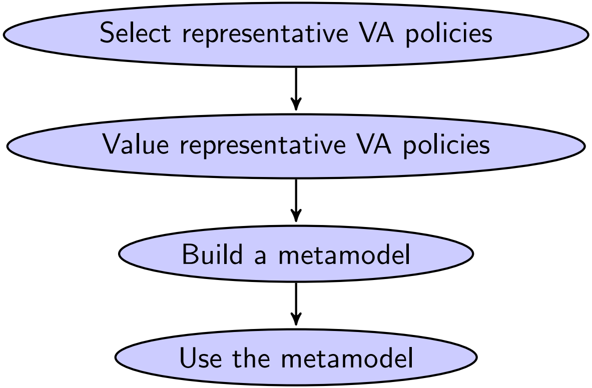
- 1,000 risk neutral scenarios
- 360 monthly time steps

$$\boxed{100,000 \times 1,000 \times 360 = 3.6 \times 10^{10}!}$$

$$\boxed{\frac{3.6 \times 10^{10} \text{ projections}}{200,000 \text{ projections/second}} = 50 \text{ hours!}}$$

Metamodeling

- A metamodel, also a surrogate model, is a model of another model.
- Metamodeling has been applied to address the computational problems arising from valuation of variable annuity portfolios: a number of work published by co-author G. Gan.
- It involves four steps:



Selecting representative policies

An important step in the metamodeling process is the selection of representative policies. Gan and Valdez (2016) compared five different experimental design methods for the GB2 regression model:

- Random sampling
- Low-discrepancy sequence
- Data clustering (hierarchical k-means)
- Latin hypercube sampling
- Conditional Latin hypercube sampling

Some metamodels proposed/examined

We have studied and proposed some metamodels for the valuation of large VA portfolios:

- Ordinary kriging
- Universal kriging
- GB2 regression model
- Rank-order kriging (quantile kriging)
- Tree-based models - joint work with Z. Quan

Kriging has its origins in geostatistics or spatial analysis. It is in some sense an interpolation method that is closely related to the idea of regression.

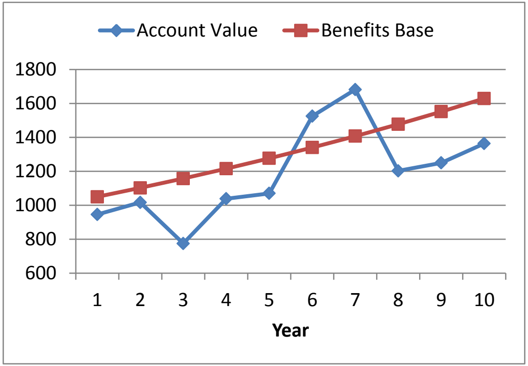
A portfolio of synthetic variable annuity policies

Feature	Value
Policyholder birth date	[1/1/1950, 1/1/1980]
Issue date	[1/1/2000, 1/1/2014]
Valuation date	1/1/2014
Maturity	[15, 30] years
Account value	[50000, 500000]
Female percent	40%
Product type	DBRP, DBRU, BBSU, etc.
Fund fee	30, 50, 60, 80, 10, 38, 45, 55, 57, 46bps for Funds 1 to 10, respectively
Base fee	200 bps
Rider fee	depends on product type
Number of funds invested	[1, 10]

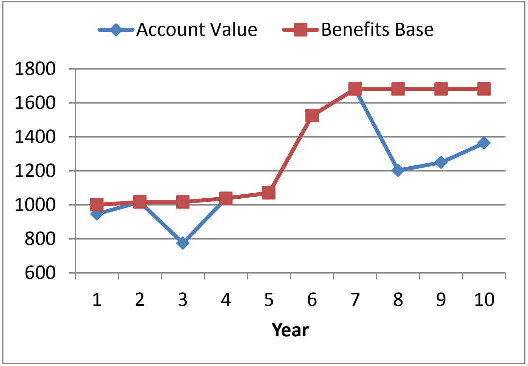
VA product types in the synthetic portfolio

Product	Description	Rider Fee
DBRP	GMDB with return of premium	0.25%
DBRU	GMDB with annual roll-up	0.35%
DBSU	GMDB with annual ratchet	0.35%
ABRP	GMAB with return of premium	0.50%
ABRU	GMAB with annual roll-up	0.60%
ABSU	GMAB with annual ratchet	0.60%
IBRP	GMIB with return of premium	0.60%
IBRU	GMIB with annual roll-up	0.70%
IBSU	GMIB with annual ratchet	0.70%
MBRP	GMMB with return of premium	0.50%
MBRU	GMMB with annual roll-up	0.60%
MBSU	GMMB with annual ratchet	0.60%
WBRP	GMWB with return of premium	0.65%
WBRU	GMWB with annual roll-up	0.75%
WBSU	GMWB with annual ratchet	0.75%
DBAB	GMDB + GMAB with annual ratchet	0.75%
DBIB	GMDB + GMIB with annual ratchet	0.85%
DBMB	GMDB + GMMB with annual ratchet	0.75%
DBWB	GMDB + GMWB with annual ratchet	0.90%

VA provides guaranteed appreciation of the benefits base

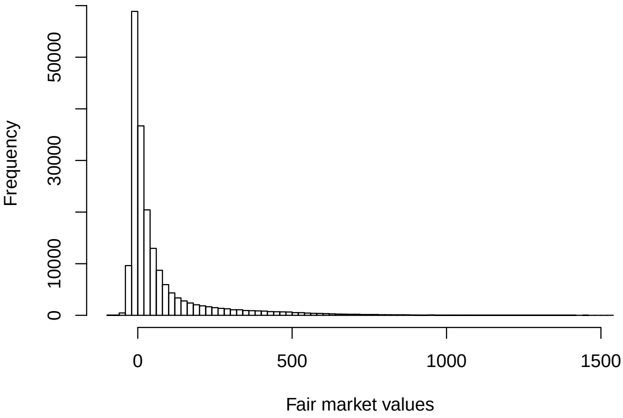


(Roll-up)



(Ratchet)

Fair market values of the guarantees



	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
`fmv`	-68.37	-5.55	64.63	11.7	64.84	1210.32

Training set - summary statistics - continuous variables

Response variables	Description	Min.	1st Q	Mean	Median	3rd Q	Max.
gmwbBalance	GMWB balance	0	0	27.8	0	0	422.26
gbAmt	Guaranteed benefit amount	51.88	183.98	323.29	306.89	437.36	920.62
FundValue1	Account value of the 1st fund	0	0	32.02	12.62	46.76	629.89
FundValue2	Account value of the 2nd fund	0	0	36.54	16.08	56.31	571.59
FundValue3	Account value of the 3rd fund	0	0	26.78	11.81	36.64	458.78
FundValue4	Account value of the 4th fund	0	0	25.8	10.48	38.29	539.36
FundValue5	Account value of the 5th fund	0	0	22.29	10.54	34.71	425.92
FundValue6	Account value of the 6th fund	0	0	37.15	19.64	53.96	654.64
FundValue7	Account value of the 7th fund	0	0	28.78	12.88	42.56	546.89
FundValue8	Account value of the 8th fund	0	0	31.27	15.59	46.24	529.57
FundValue9	Account value of the 9th fund	0	0	31.93	13.9	45.17	599.44
FundValue10	Account value of the 10th fund	0	0	32.6	13.86	45.09	510.43
age	Age of the policyholder	34.52	42.86	50.29	51.36	57.21	64.46
ttm	Time to maturity in years	0.75	10.09	14.61	14.6	19.12	27.52

Tree-based models

Quan, Gan and Valdez (2019) compared the prediction performance of various tree-based models:

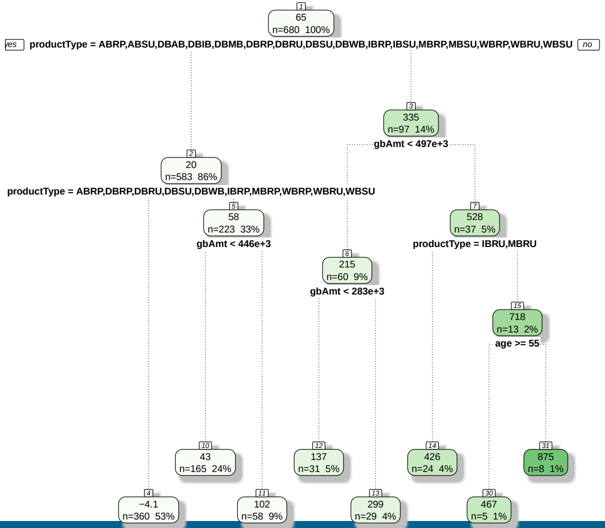
- Classification and Regression Trees (CART)
  - pruned by introducing penalty
- Ensemble methods: aggregate several regression trees to improve prediction accuracy
  - Bagging and random forests
  - Gradient boosting
- Unbiased recursive partitioning:
  - Conditional inference trees
  - Conditional random forests

Unbiased recursive partitioning

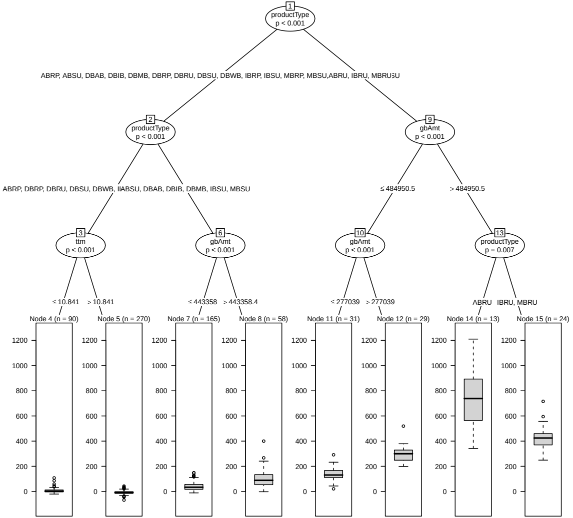
CART algorithms employ what is called recursive binary partitioning, which uses greedy search causing some drawbacks:

- Overfitting
  - Use a pruning process by applying cross-validation
- Bias in variable selection
  - Especially true when the explanatory variables present many possible splits or have missing values
  - Hothorn, et al. (2006) introduced conditional inference trees based on a partitioning of a statistic that is used to measure the association between the response and the explanatory variables.

A regression tree



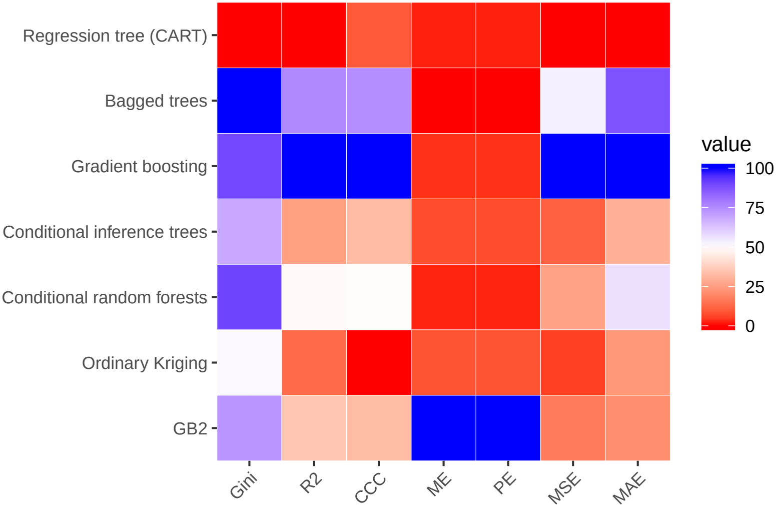
A conditional inference tree



Prediction accuracy of various models

Model	$Gini$	R^2	CCC	ME	PE	MSE	MAE
Regression tree (CART)	0.786	0.845	0.917	1.678	-0.025	3278.578	31.421
Bagged trees	0.842	0.918	0.954	2.213	-0.033	1720.725	20.334
Gradient boosting	0.836	0.942	0.969	1.311	-0.019	1214.899	19.341
Conditional inference trees	0.824	0.869	0.930	0.905	-0.013	2754.853	26.536
Conditional random forests	0.836	0.892	0.940	1.596	-0.024	2273.385	23.219
Ordinary Kriging	0.815	0.857	0.912	-0.812	0.012	3006.192	27.429
GB2	0.827	0.879	0.930	0.106	-0.002	2554.246	27.772

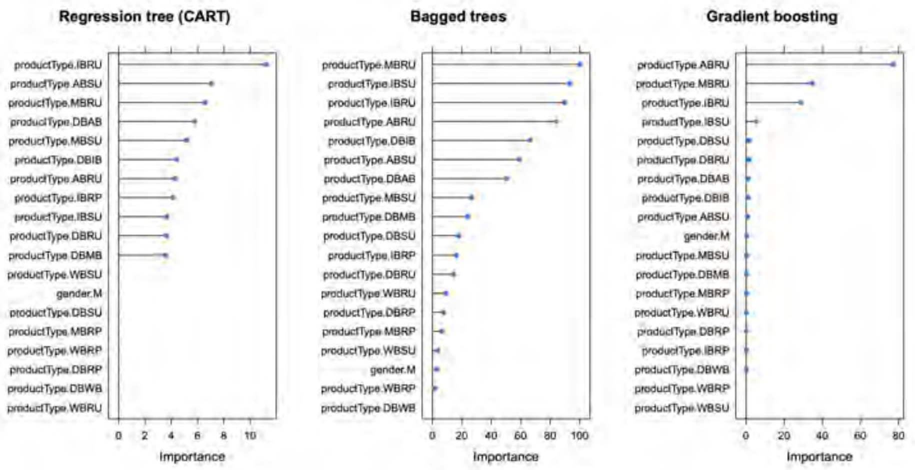
A heatmap of model performance



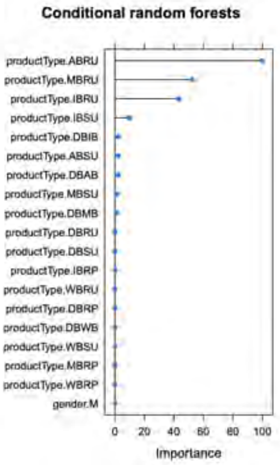
Computational efficiency

Model	Computation Time
Regression tree (CART)	0.13 secs
Bagged trees	2.70 secs
Gradient boosting	4.69 secs
Conditional inference trees	0.25 secs
Conditional random forests	1214.72 secs
Ordinary Kriging	277.49 secs
GB2	23.44 secs

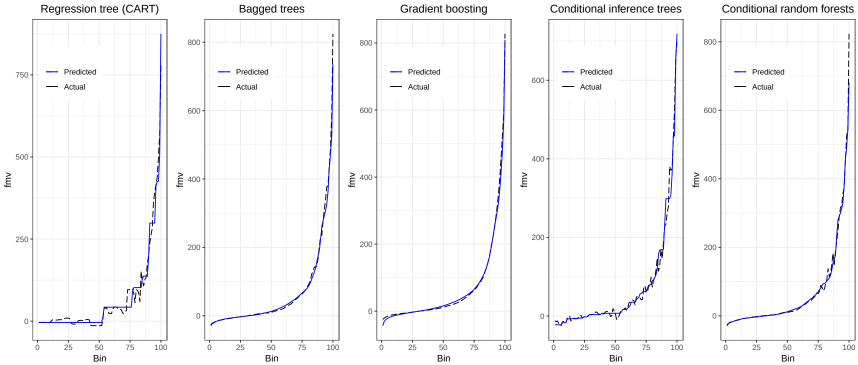
Variable importance for tree-based models



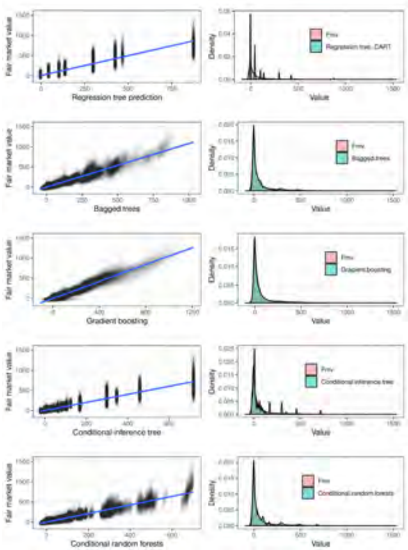
Variable importance for tree-based models



Lift curve plots - performance visualization



Prediction and observed fair market values



Concluding remarks

We explore tree-based models and their extensions in developing metamodels for predicting fair market values. Besides computational efficiency and predictive accuracy, they have several advantages as an alternative predictive tool:

- Tree-based models are considered as nonparametric models that do not require distribution assumptions.
- Tree-based models can perform variable selection by assessing the relative importance.
- Tree-based models, especially with single smaller-sized trees, are straightforward to interpret by a visualization of the tree structure. This visualization was illustrated both in the case of regression tree and conditional inference tree.
- When compared to other metamodels for prediction purposes, tree-based models require less data preparation as they preserve the original scale to be more interpretable.

Metamodeling book

Appendix: Validation measures

Validation measure	Description	Interpretation
Gini Index	$Gini = 1 - \frac{2}{N-1}\left(N - \frac{\sum_{i=1}^{N} i\tilde{y}_i}{\sum_{i=1}^{N} \tilde{y}_i}\right)$ where $\tilde{y}$ is the corresponding to y after ranking the corresponding predicted values $\widehat{y}$.	Higher Gini is better.
Coefficient of Determination	$R^2 = 1 - \frac{\sum_{i=1}^{N} (\widehat{y}_i - y_i)^2}{\sum_{i=1}^{N} \left(y_i - \frac{1}{n}\sum_{i=1}^{n} y_i\right)^2}$ where $\widehat{y}$ is predicted values.	Higher R^2 is better.
Concordance Correlation Coefficient	$CCC = \frac{2\rho\sigma_{\widehat{y}_i}\sigma_{y_i}}{\sigma^2_{\widehat{y}_i} + \sigma^2_{y_i} + (\mu_{\widehat{y}_i} - \mu_{y_i})^2}$ where $\mu_{\widehat{y}_i}$ and μ_{y_i} are the means $\sigma^2_{\widehat{y}_i}$ and $\sigma^2_{y_i}$ are the variances ρ is the correlation coefficient	Higher CCC is better.
Mean Error	$ME = \frac{1}{N}\sum_{i=1}^{N}(\widehat{y}_i - y_i)$	Lower $\|ME\|$ is better.
Percentage Error	$PE = \frac{\sum_{i=1}^{N} \widehat{y}_i - \sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} y_i}$	Lower $\|PE\|$ is better.
Mean Squared Error	$MSE = \frac{1}{N}\sum_{i=1}^{N}(\widehat{y}_i - y_i)^2$	Lower MSE is better
Mean Absolute Error	$MAE = \frac{1}{N}\sum_{i=1}^{N}\|\widehat{y}_i - y_i\|$	Lower MAE is better.

Appendix: Tuning hyperparameters

R package	Description
`rpart`	Classification and regression tree (CART)
cp	complexity parameter
minsplit	minimum number of observations in a node in order to be considered for splitting
maxdepth	maximum depth of any node of the final tree
`randomForest`	Bagging and Random Forests
mtry	number of explanatory variables randomly sampled as candidates at each split
nodesize	minimum number of observations in the terminal nodes
ntree	number of trees to grow/bootstrap samples
`gbm`	Gradient boosting
n.trees	number of trees to fit/iterations/basis functions in the additive expansion
interaction.depth	maximum depth of variable interactions(1 implies an additive model, 2 means a model with up to 2-way interactions)
n.minobsinnode	minimum number of observations in the terminal nodes
shrinkage	shrinkage parameter(learning rate or step-size reduction)
`party/partykit`	Conditional inference trees
teststat	type of the test statistic to be applied for variable selection
splitstat	type of the test statistic to be applied for split point selection
testtype	the way to compute the distribution of the test statistic
alpha	significance level for variable selection
minsplit	minimum sum of weights in a node in order to be considered for splitting
`party/partykit`	Conditional random forests
mtry	number of explanatory variables randomly sampled as candidates at each split
ntree	number of trees to grow/bootstrap samples

References

- Breiman, L., et al. (1984). *Classification and Regression Trees*. Taylor & Francis Group, LLC: Boca Raton, FL.
- Gan, G. and Valdez, E.A. (2019). *Metamodeling for Variable Annuities*. CRC Press: Boca Raton, FL.
- Gan, G. and Valdez, E.A. (2017). Valuation of large variable annuity portfolios: Monte Carlo simulation and synthetic datasets. *Dependence Modeling*. 5:354-374.
- Gan, G. and Valdez, E.A. (2018). Regression modeling for the valuation of large variable annuity portfolios. *North American Actuarial Journal*. 22(1):40-54.
- Hothorn, T., Hornik, K. and Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*. 15(3):651-674.
- Loh, W.-Y. (2014). Fifty years of classification and regression trees. *International Statistical Review*. 82(3):329-348.
- Quan, Z., Gan, G. and Valdez, E.A. (2019). Tree-based models for variable annuity valuation: Parameter tuning and empirical analysis. Submitted for publication.