2018 Predictive Analytics Symposium

Session 26: AP - Assessing Credibility of Predictive Model

Assessing credibility of predictive models

Moderator: Stuart Klugman, FSA, CERA, PhD

Presenter: Matthias Kullowatz, ASA, MAAA, MS

September 19, 2019

SOCIETY OF ACTUARIES®

SOA 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.

Introduction

SOCIETY OF ACTUARIES®

Agenda

1. Pleased to meet you
2. Define key terms
3. Predictive model validation
4. Useful statistical techniques:
   a. Linear mixed effects models
   b. Elastic net regularized (“penalized”) models
   c. Limited fluctuation application to GLMs (if time!)
5. Questions and discussion
6. Github repository:

   https://github.com/milliman/SOA_PAS_CrediblePredictiveModels

Definitions

SOCIETY OF ACTUARIES®

Credibility

ASOP 25:

- *A measure of the predictive value in a given application that the actuary attaches to a particular set of data (predictive is used here in the **statistical sense** and not in the sense of predicting the future).*
- *In [predictive models], credibility can be estimated based on the **statistical significance** of parameter estimates, model performance on a holdout data set, or the consistency of either of these measures over time.*

"Statistical sense" and "statistical significance" suggest a focus on data ***quantity***

Credibility

ASOP 25:

- *A measure of the predictive value in a given application that the actuary attaches to a particular set of data (predictive is used here in the statistical sense and not in the sense of predicting the future).*
- *In [predictive models], credibility can be estimated based on the statistical significance of parameter estimates,* ***model performance on a holdout data set****, or the consistency of either of these measures over time.*

Other credibility considerations

Data integrity

- Large set of bad data worse than small set of bad data
- See: ASOP 23

Predictive model appropriateness

- Various methods will lead to various sets of “credible” predictions. Is one method most appropriate?

Bias-variance tradeoff

- Minimize this: $E\left[\left(y-\hat{f}(x)\right)^2\right]$

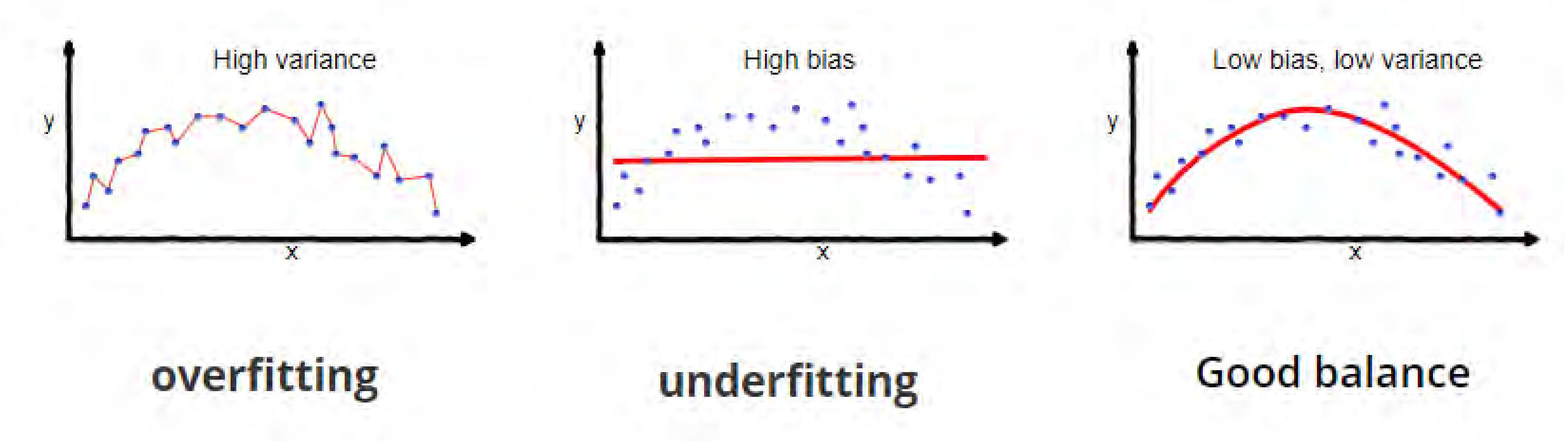
$$= \left(E\left[\hat{f}(x)-f(x)\right]\right)^2 + \left(E\left[\hat{f}(x)^2\right] - E\left[\hat{f}(x)\right]^2\right) + Var(y)$$

$$= bias^2 + (model)\ variance + random\ error$$

- A model complexity tradeoff between over- and under-fitting…between bias and variance.

Bias-variance breakdown

- The goal is to find the sweet spot; balance bias and variance so that overall error is low
  - Separate the signal from the noise effectively



https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229

Tabular model

- Data are segmented by a few dimensions, average outcomes are calculated in each segment
- Note that this *is* a predictive model
- Bias-variance tradeoff: Low-bias, high-variance

Attained age	Sex	Smoker	q
…			
65	M	S	0.010
65	M	N	0.005
65	F	S	0.008
65	F	N	0.004
66	M	S	0.011
…			

Predictive model*

- A model with the ability to consolidate segments by identifying patterns between covariates and the outcome variable
- Can control model complexity more easily; can control bias-variance tradeoff
- Including (but not limited to):
    - Linear regression
        - Regularization and mixed effects
    - Bayesian models
    - Decision trees (and ensembles thereof)
    - Support vector machines
    - Neural networks

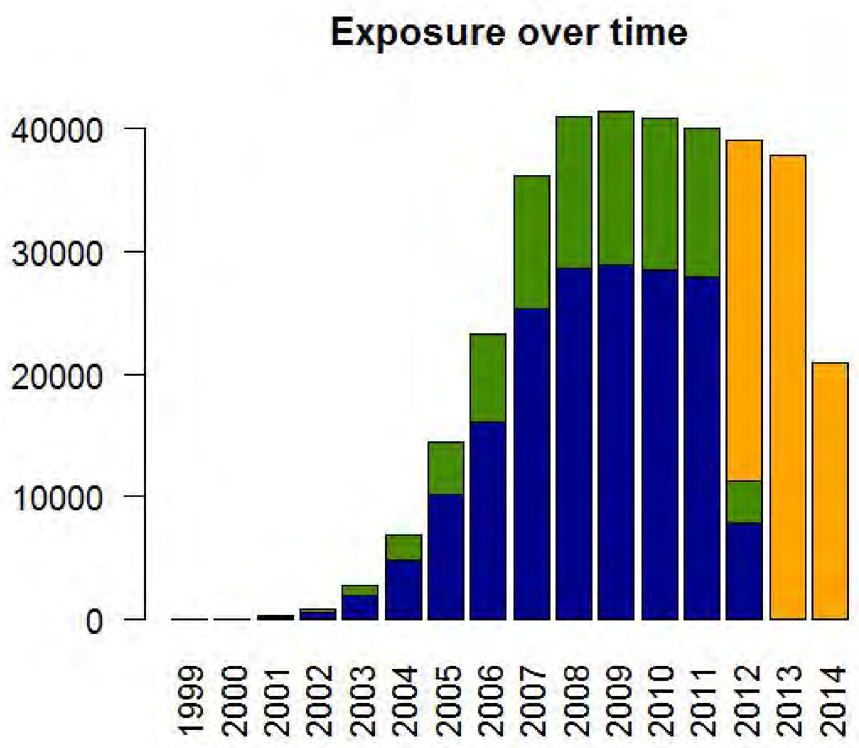
A credible predictive model

1. The data are adequately cleaned and reviewed for reasonableness
2. The data are plentiful enough to generate confident predictions across a useful range of relevant dimensions
3. The predictive modeling method used is able to optimize the bias-variance tradeoff
   - Validation: use of holdout datasets to test goodness of fit
   - Consistency of parameter estimates and validation over time

Model validation

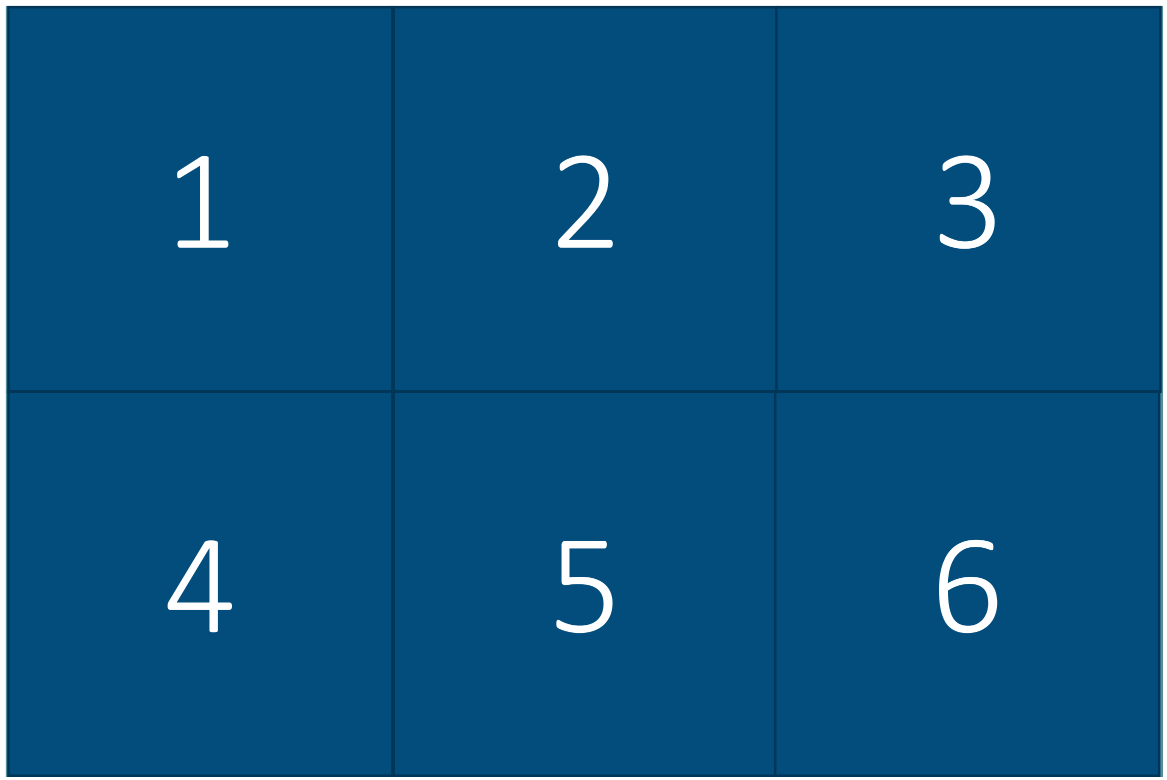
SOCIETY OF ACTUARIES®

Training versus holdout data



Cross-validation

- Useful for smaller datasets

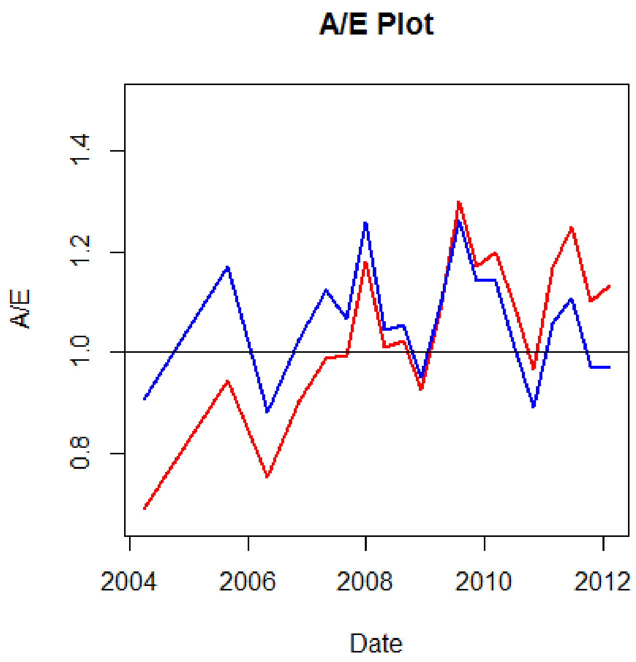


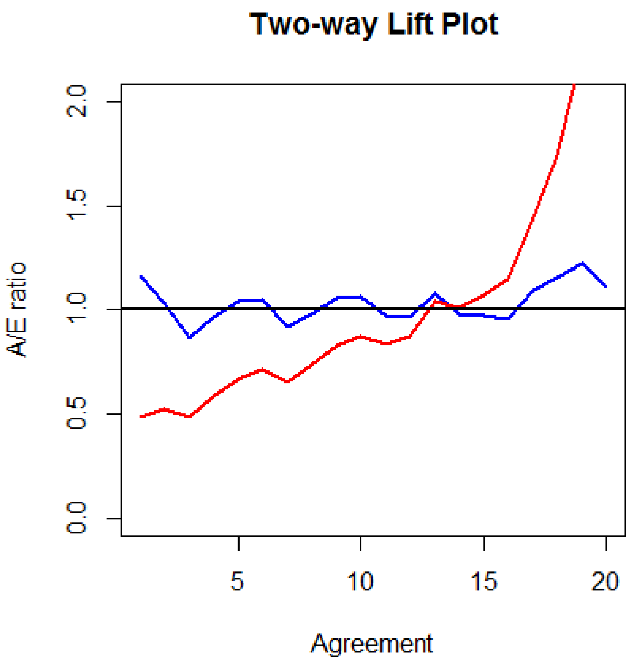
Model fit

- R^2
- Log-likelihood/AIC/BIC
- Actual-to-expected plots
  - Two-way lift charts
- Area under the ROC curve

Model comparison: Lift charts

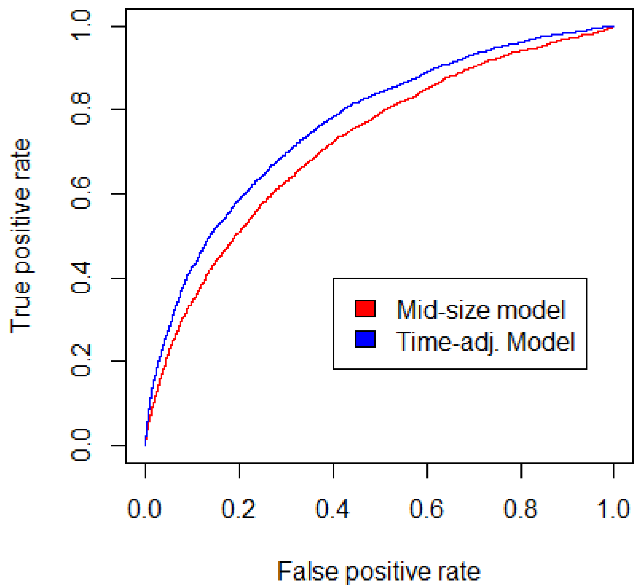
- Actual-to-expected
- Two-way lift





Area under the curve (AUC)

- The curve here is the relationship of the true positive rate and false positive rate as the positive/negative threshold moves from 0 to 1



Linear mixed effects

SOCIETY OF ACTUARIES®

Linear mixed effects models

- Model parameters can be fit as a random effects, which allows for some family-wise shrinkage
- Coefficients that are fit based on “small sample sizes” are shrunk toward zero
  - Similar to regularization in its goal
- “lme4” package in R has a relatively efficient algorithm for such fits

Linear mixed effects model (LMM)

Case study:

- A company uses many distinct distributors, but not a lot of exposure for many of the distributors. It would like to generate “credible” surrender predictions for policyholders, *taking distributor into account*.
- Many distributors will have limited exposure in the datasets
- GLMs struggle with fixed effects where either 0% or 100% of the observations recorded an event

LMM data and model

Data:

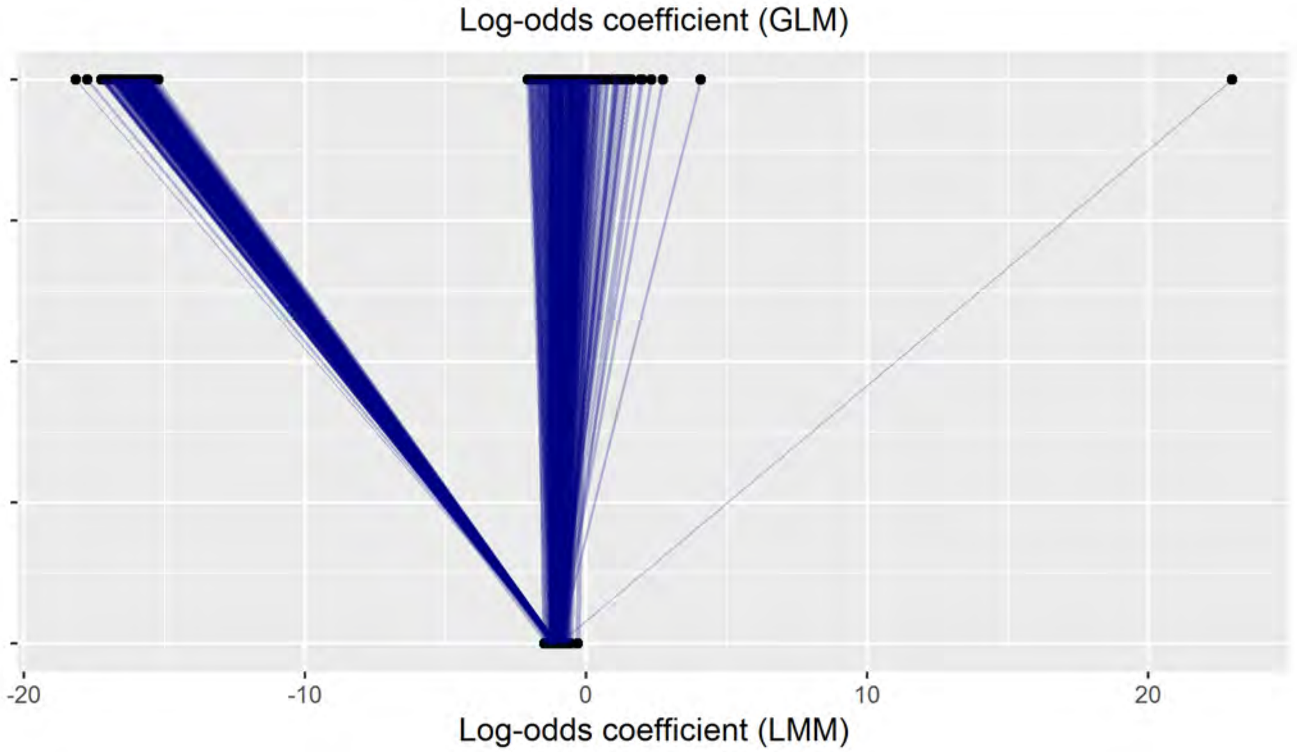
- 50,000 observations from 10,000 policyholders, undersampled to produce a 2% quarterly surrender rate.

Model:

- Probability of surrender (Surr) is a function of **moneyness** of the guarantee (ITM), **duration** (q), the **surrender charge phase** (IN, OUT), and **distributor** (DistCode)
- GLM
  - Surr ~ IN + Dur_IN + ITM:Dur_IN + Dur_OUT + ITM:Dur_OUT + DistCode
- LMM
  - Surr ~ IN + Dur_IN + ITM:Dur_IN + Dur_OUT + ITM:Dur_OUT + **(1 | DistCode)**

LMM continued

Fixed distributor effects from the GLM are shrunk to more credible random effects in the LMM:



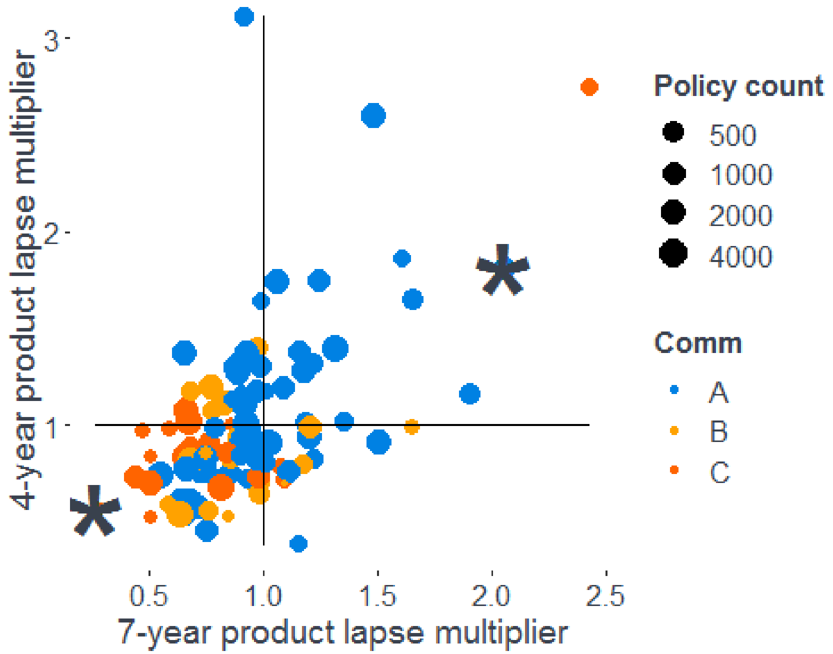
LMM interpretation

- Coefficient comparison: Even with shrinkage, we get differentiation of distributors
  - More than 3x surrender rates from maximum effect to minimum
  - More than 1.5x surrender rates across middle 95%
- Model comparison:
  - LMM fits much better than GLM on holdout dataset (not including DistCode)

LMM credibility

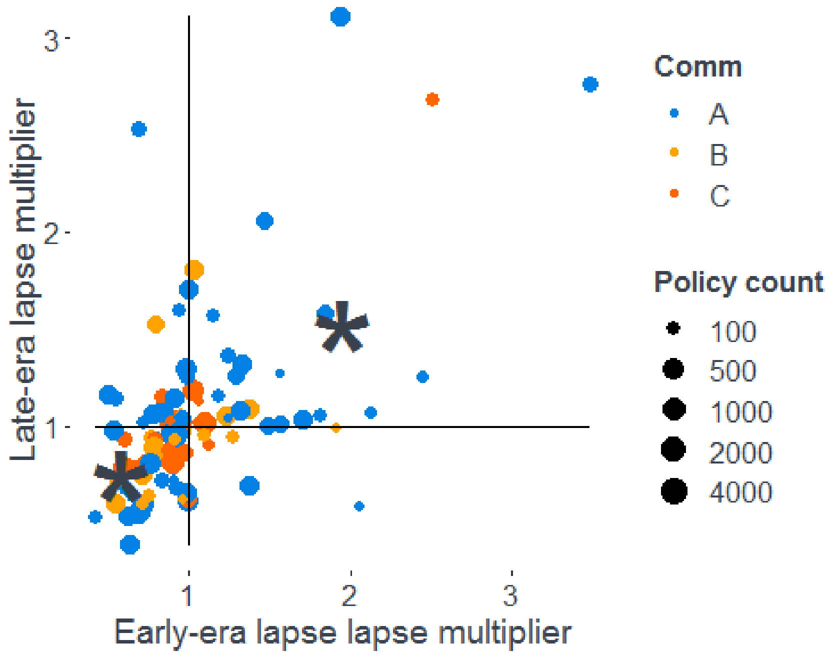
- We were able to arrive at more credible surrender predictions by distributor…
- …and more credible estimates of the effects of each distributor
- What makes the predictions credible overall?
  - The random effects process
  - Performance on a holdout dataset, and the consistency thereof over time (ASOP 25)
- What can we do with these models?!

Consistency across product



- We observe consistent effects by distributor across products (4- and 7-year SC products).
- These effects can help explain periodic variations in observed lapse behavior.
- This framework can identify anti-selective commission elections and quantify the cost of these behaviors.

Consistency across issue era



- We observe consistent distributor effects across issue eras.
- This framework can provide insights to inform policy recommendations around financial advice and sales/marketing practices.
- Example: Explore effects of proposed DOL rules by strategically setting issue era split dates; if rules were implemented

Regularization with offset

Using offset as “credibility target”

What is an offset?

- A covariate that is constrained to have a coefficient of 1
- Like a null hypothesis

How is it implemented?

- Input as a vector in most R model functions
- **Should be in units of the prediction**

What can it do for us?

- Input some known credible assumptions as an offset to serve as a credibility target

Regularization with offset

- Additional constraints (“penalties”) on the model coefficients
- Set the offset to some “credibility target”
  - The offset effectively forces the model to identify regions where the data stray from the null hypothesis
- The coefficient constraints serve as a credibility weighting method
- Where data are scarce, the model will be weighted heavily toward the offset

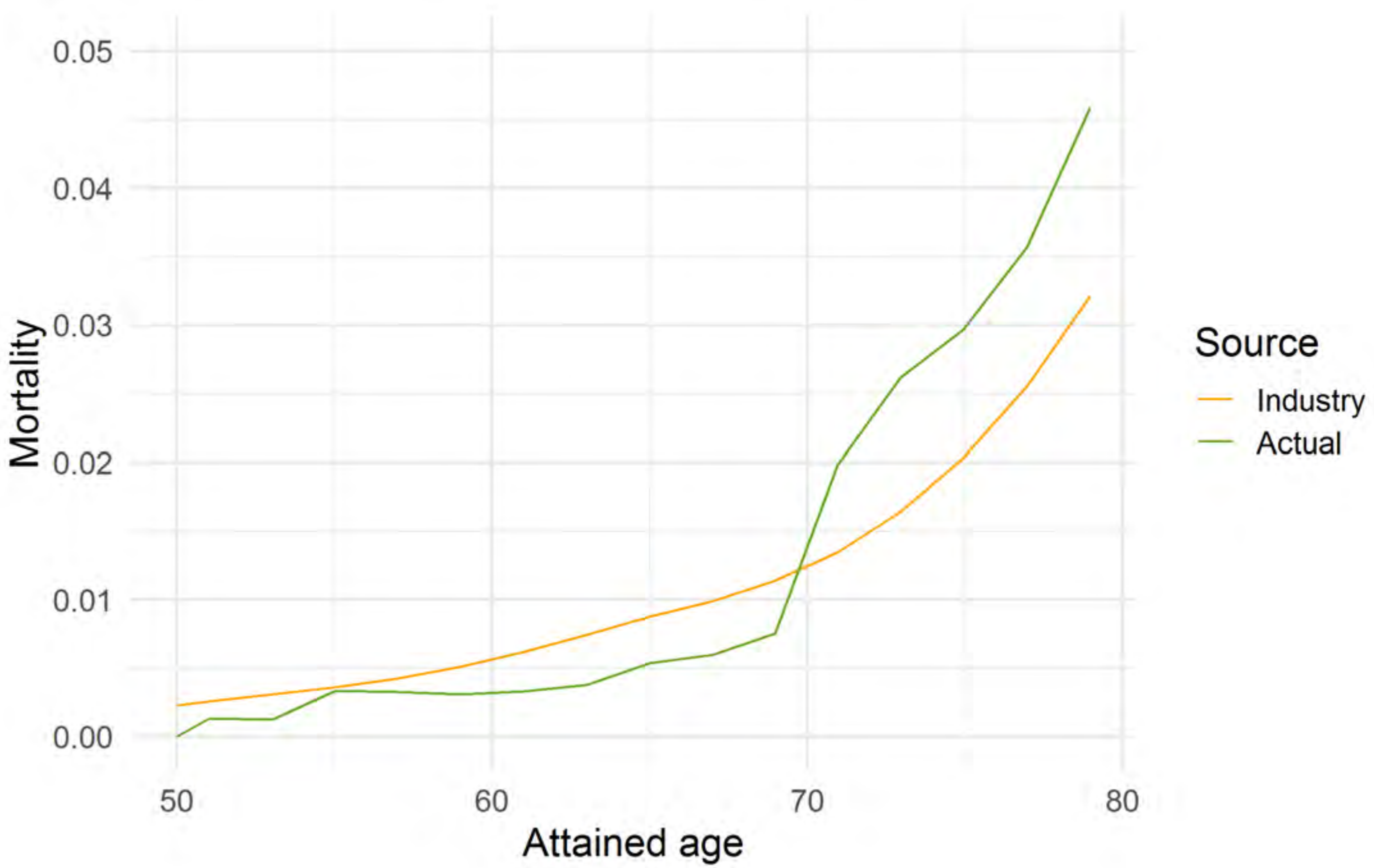
Logistic regression offset

- $\ln\left(\frac{\hat{q}}{1-\hat{q}}\right) = \widehat{\beta_0} + \widehat{\beta_1}x_1 + \cdots + \widehat{\beta_p}x_p + 1 \cdot offset$
- Observation *i* has prescribed mortality of $q_i = 0.01$
- $\ln\left(\frac{q_i}{1-q_i}\right) = \ln\left(\frac{0.01}{0.99}\right) = -4.595$

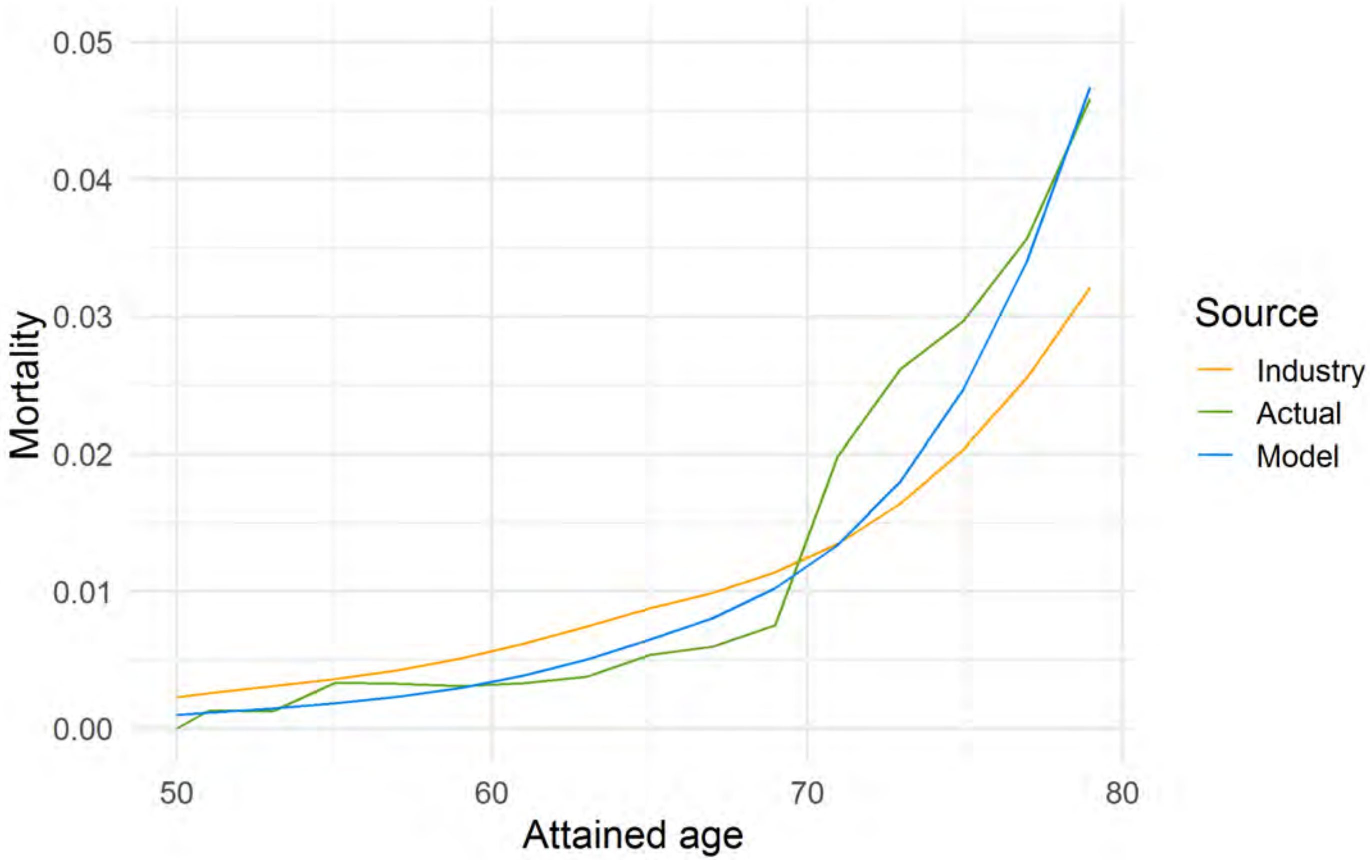
Regularization details

- Penalties:
  - GLM: maximize(log-likelihood)
  - Lasso: maximize(log-likelihood - $\boldsymbol{\lambda} \sum \left|\boldsymbol{\beta_j}\right|$)
  - Ridge: maximize(log-likelihood - $\boldsymbol{\lambda} \sum \boldsymbol{\beta_j^2}$)
- Requires standardized covariates because the value of the coefficient is part of the penalty
- “glmnet” package in R

Case study: VA mortality



Case study: VA mortality



Limited fluctuation credibility

SOCIETY OF ACTUARIES®

Limited fluctuation credibility (LFC)

- More credible when the probability of proportionally small errors (k) is large (1 – α)
  - $P(|\bar{X} - \mu| < k\mu) > 1 - \alpha$
- For an observed cell, probability distribution typically comes from Central Limit Theorem
- We often look for at least 1,082 events in a cell for "full credibility." Where does 1,082 come from?

1,082 derivation (proportions)

- Recall: $\mu = q;\ \sigma^2 = \frac{q(1-q)}{n}$
- Work through some algebra and probability theory:

1. $P(|\hat{q} - q| < kq) > 1 - \alpha$
2. $P\left(\left|\frac{\hat{q}-q}{\sqrt{\frac{q(1-q)}{n}}}\right| < \frac{kq}{\sqrt{\frac{q(1-q)}{n}}}\right) > 1 - \alpha$
3. $\boldsymbol{P\left(|Z| < \frac{k\sqrt{nq}}{\sqrt{1-q}}\right) > 1 - \alpha}$
4. $\boldsymbol{nq > \left(\frac{Z_{\frac{\alpha}{2}}}{k}\right)^2 \cdot (1 - q)}$
5. *Choose $k = 5\%$ and $1 - \alpha = 90\%$*
6. $nq > \left(\frac{1.645}{0.05}\right)^2 (1 - q) = 1{,}082.4(1 - q)$ and 1 – q is typically very close to 1

SOCIETY OF ACTUARIES

LFC using confidence intervals

$$3.\ P\left(|Z| < \frac{k\sqrt{nq}}{\sqrt{1-q}}\right) > 1 - \alpha \ \Rightarrow\ 4.\ \boldsymbol{Z_{\frac{\alpha}{2}}} \cdot \sqrt{\frac{\boldsymbol{\hat{q}(1-\hat{q})}}{\boldsymbol{n}}} \le k \cdot \hat{q}$$

Width of confidence interval: $\mathbf{2} \cdot \boldsymbol{Z_{\frac{\alpha}{2}}} \cdot \sqrt{\frac{\boldsymbol{\hat{q}(1-\hat{q})}}{\boldsymbol{n}}}$

Width of error tolerance interval: $\mathbf{2} \cdot \boldsymbol{k} \cdot \boldsymbol{\hat{q}}$

LFC with a GLM

- GLM coefficients (and thus predictions) have approximate normal distributions, with variances derived from the variance-covariance matrix
- Compare GLM prediction confidence interval to an error tolerance interval to determine full credibility
- All can be done in base R

https://www.soa.org/globalassets/assets/library/newsletters/predictive-analytics-and-futurism/2017/december/2017-predictive-analytics-newsletter-issue-16.pdf

LFC with a GLM pros and cons

Pros

- Follows from commonly used credibility method
- GLMs are interpretable and commonly used

Cons

- No obvious weighting scheme for predictions with less than full credibility
  - $Weight = "Z" = \sqrt{\frac{n}{n^*}}$

Lightning round: things to consider

SOCIETY OF ACTUARIES®

Things to consider

1. If underlying distribution changes over time, are you relying on a consistency that doesn't actually exist?
   - E.g. rising interest rate scenario
   - Mortality: improvement, cure shocks, etc.
2. Weighting recent data vs. distant past data
3. Thresholds for data cleanliness
4. What proportion of deriving credible estimates should be qualitative?

Bayesian model (MCMC)

- Modeler declares model parameters, their prior distributions, and their likelihood functions
- “Coefficients” are represented by posterior distributions, derived from sampling the data and Bayes’ theorem
- “rstan” package in R

“Bayesian inference in Machine Learning”

https://www.soa.org/Library/Newsletters/Predictive-Analytics-and-Futurism/2017/june/2017-predictive-analytics-newsletter-issue-15.pdf

Bayesian model pros and cons

Pros

- Flexible model specification
- Priors are an intuitive form of “credible targets”
- Produces full distributions of parameters (coefficients)

Cons

- Computationally intensive
- Steeper learning curve: specifying the model, priors, and likelihood functions can be an involved process

Questions

Thank you!

SOCIETY OF ACTUARIES®