ILEC MACHINE LEARNING: FROM BASIC TO EXTRA

PHILIP L. ADAMS, FSA, CERA, MAAA
SENIOR EXPERIENCE STUDIES ACTUARY
SOCIETY OF ACTUARIES RESEARCH INSTITUTE

SOUTHEASTERN ACTUARIES CONFERENCE
ATLANTA, GA | NOVEMBER 20, 2024

BUILDING A PREDICTIVE ANALYTICS FRAMEWORK

BUILDING TOWARD A FRAMEWORK

Aims

Develop a framework for model building and insight discovery

Applicable to outputs of experience studies systems

Focusing theme for this project: mortality differences by product

Audience

Experience analytics practitioners looking for workflow to enhance analysis

Fresh recipients of the predictive analytics certificate looking for an application of what they learned

Process

Prepare and supply your data

Specify inputs and outputs

Set relevant parameters

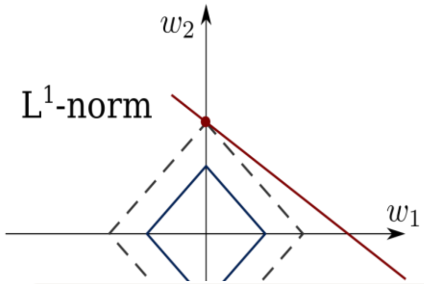
GIST OF THE FRAMEWORK

Poisson Main Effects GLM

For initial explorations

Gradient boosting decision tree (GBDT)

For important features and 2-way interactions



Elastic Net GLM

For bringing it all together, *explainably and credibly*

GLM FINDINGS

AMOUNT-BASED, 15VBT OFFSET

		Ratios of Ratios		
		Perm-to-Term	Perm-to-UL/VL	Perm-to-Other
A/E 2015VBT by Amount		115.6%	108.1%	114.6%
GLM Fitted Factors		115.1%	91.2%	87.5%
Ratio of Ratios		100.4%	118.4%	131.0%
Ratio of Weighted Average Factors*	Duration	100.0%	100.9%	97.7%
	Face Amount	107.8%	107.3%	108.4%
	Gender	99.9%	100.1%	100.0%
	Issue Age	98.8%	103.9%	101.1%
	Issue Year	105.3%	104.1%	107.7%
	Level Term Period	79.6%	100.0%	100.0%
	Underwriting	111.9%	101.0%	103.5%
	Product of Ratios	99.9%	118.5%	119.3%

Strong distributional differences, modeling approach must accommodate this

Application of Brian Holland's paper on understanding GLM features

Ratio-of-ratios analysis: how does the weighted average of a GLM component change when moving from subset to subset?

Deviation from 100% points to different underlying prevalence, and potentially need for interactions (interaction variables, partitioning)

GRADIENT-BOOSTING DECISION TREE *(GBDT)*

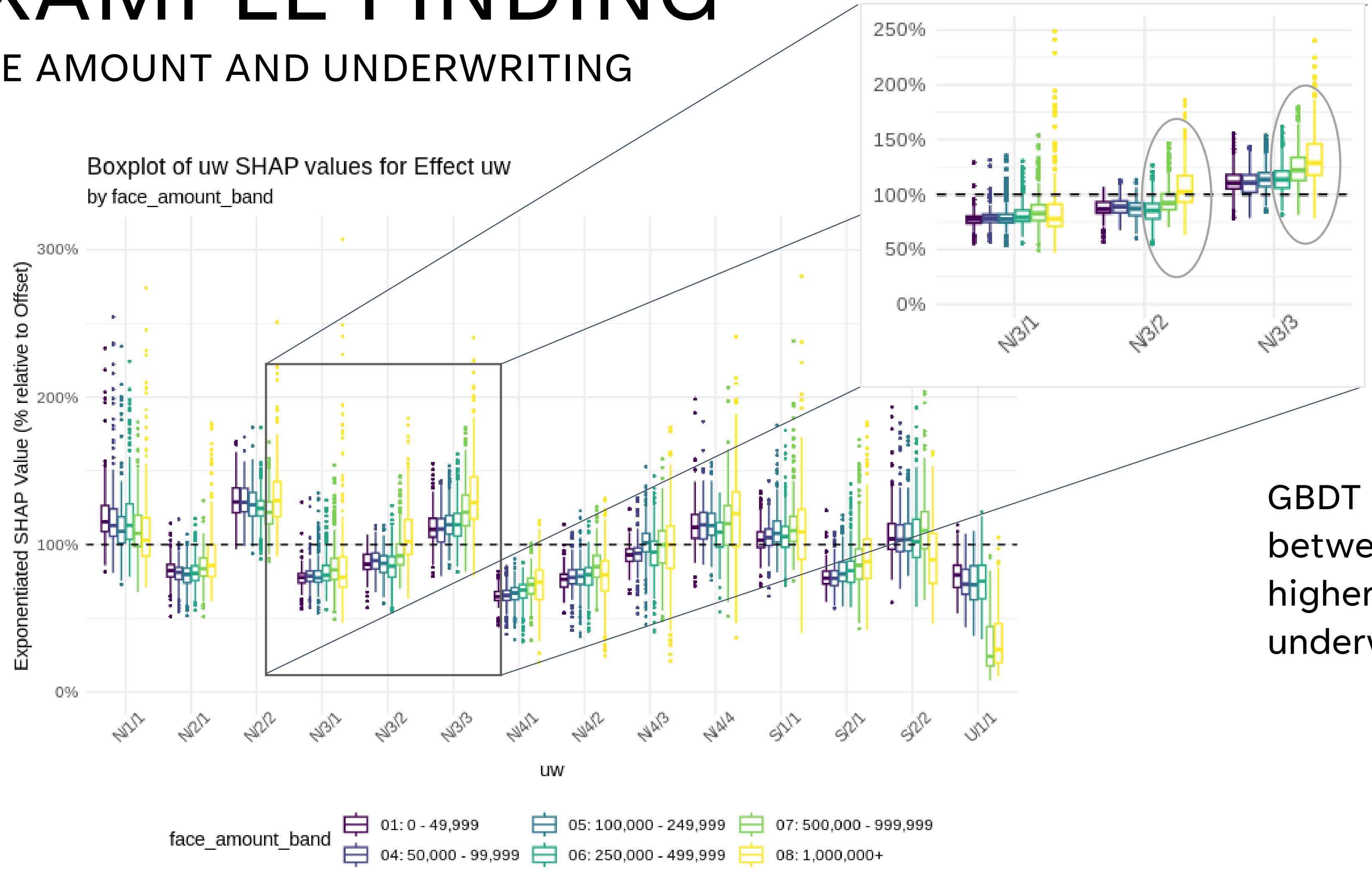
- **Purpose: split adjustments into higher and lower subj. to constraints**
- Intuition:
  - Generate a shallow tree of adjustments: splitting adjustments into higher and lower as experience indicates from one initial overall adjustment
  - Apply these adjustments, and fit another shallow tree against residual variation
  - Do this until no more variation to adjust for
  - Result is a stack of recursively fitted, shallow decision trees
- In practice:
  - Tree fitting starts where data are heaviest, and works its way down to where data are lightest
  - Here using "LightGBM" implementation, splitting leaves as needed
  - Datasets can require 100s to thousands of tree fits

EXPLAINING GBDT WITH SHAP VALUES

- **Purpose: show marginal contribution of feature/factor/variable to outcome**
- Intuition: when bringing in a variable, measure how the that variable changes model
- High Level Theory:
  - For all orders of adding variables, average the change in each model estimate.
  - That can be so many: for n features, there are 2^n different models!
  - Practical solution – this change itself is estimated.
- In Practice for GBDT
  - Each split in a tree changes prediction, so accumulate those for all variables within and across trees
  - Example: first split on face amount increases prediction 11%, then next split decreases for gender by 3%, then next split on face amount increases prediction by 13%; final tally for gender is 3%, face amount band is 25.4%

EXAMPLE FINDING

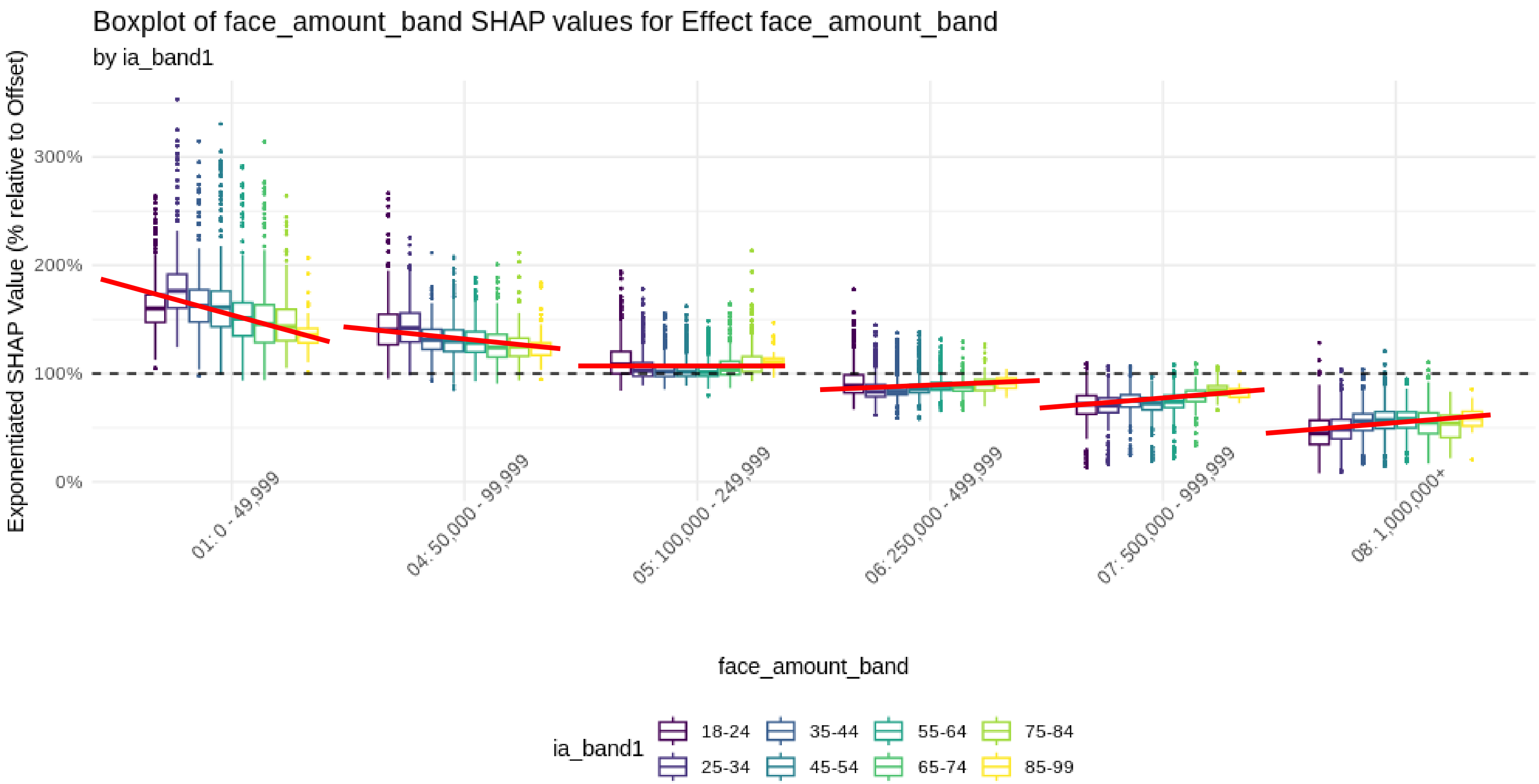
FACE AMOUNT AND UNDERWRITING



GBDT detects interaction between face amount and higher face 3-class underwriting

EXAMPLE FINDINGS

FACE AMOUNT AND ISSUE AGE



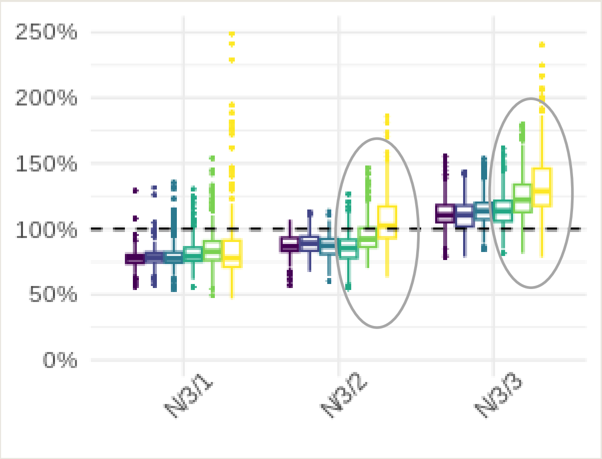
- Issue age slopes vary by face amount, or
- Face amount band spread decreases with issue age

ELASTIC NET

- Purpose: fit a regression model, dealing with offsetting or negligible factors through regularization (penalizing size of coefficients)
- Intuition: limit size of coefficients, balancing completely excluding small coefficients (L1 regularization) and limiting size of potentially offsetting coefficients (L2 regularization)
- Example: duration and attained age move together, so avoid large offsetting coefficients with opposite signs

ELASTIC NET FINDINGS

- Of 28 possible interaction terms, 23 survived to the end
- Notable interactions relating to product
  - Underwriting: average spread of UW factors higher for perm and term, lower for UL/VL overall
  - Face amount band: lower face amounts for perm tend to have better factors than for term (echoing socioeconomic influences?)
- Factor for 3-class residual standard for face amount 1M+: 114.7%



PATHS NOT TAKEN: FUTURE POSSIBILITIES

- Things we could do differently
  - Adaptive LASSO (easy)
  - LASSO Confidence Intervals (hard)
  - Handling of continuous covariates (easy)
- Things we might have tried
  - Bayesian methods (STAN or INLA)
  - Deep learning methods

LESSONS LEARNED

GitHub

Technical learning curve

Generalizing data science

Reasoning is hard to automate

Your problem will differ

Computing burden

Memory

Computing

Platform

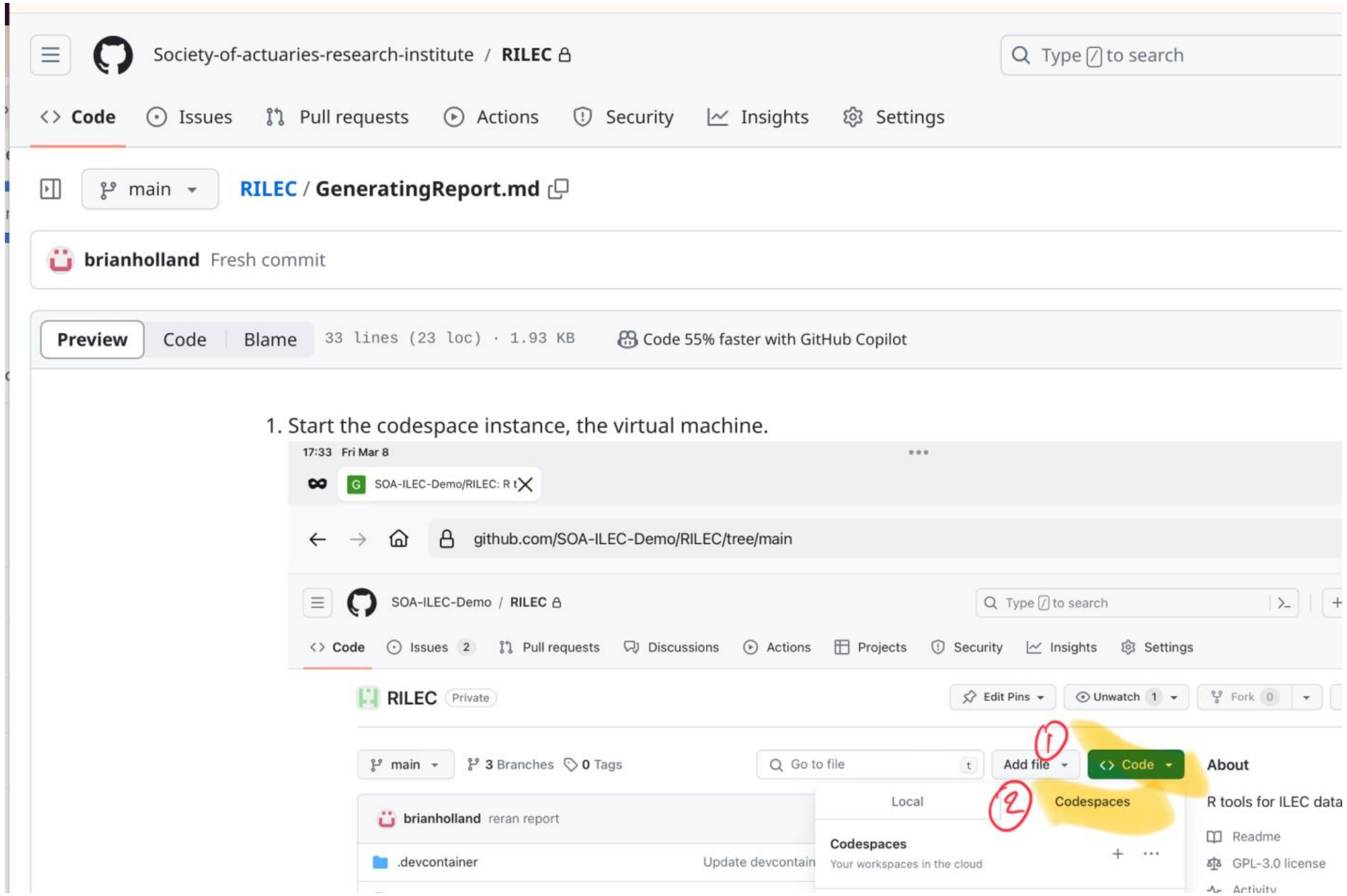
TECH STACK – RUN IT YOURSELF TOO

Tools used:

- Git/GitHub: change tracking, task assignment, peer reviews
- CodeSpaces: common virtual environment atop GitHub (on your own system or Github's)
- VSCode: one editor that works with GitHub and CodeSpaces
- Rstudio: interactive use of R, running scripts, viewing results
- *Reasonably priced*
- https://github.com/Society-of-actuaries-research-institute/RILEC

RSTUDIO IN CODESPACES

- Generate interactive HTML report in RStudio
- Instructions in file *GeneratingReport.md ->*



STRAY THOUGHTS

- Limitations
  - Framework only works on categorical predictors end-to-end (continuous could be added)
  - Does not tell you how to partition data, nor do any partitioning for you
- Where to find it
  - https://github.com/Society-of-actuaries-research-institute/RILEC
- Connections between credibility and penalization
  - *Applying Credibility in Penalized Regression, 2023, Akur8, CAS*
  - *Applying Penalized Credibility as a Credibility Procedure, 2023, SOA Webcast*
  - *A discussion on credibility and penalized regression, ..., 2015, Hugh Miller, ASTIN*

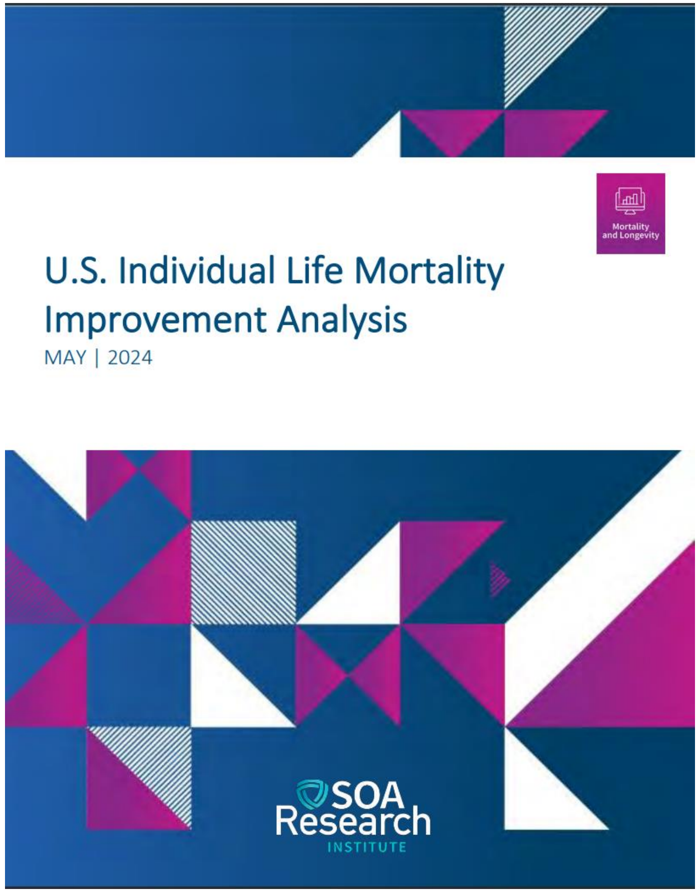
THANKS

- The team: Philip Adams, Michael Niemerg, Haofeng Yu, Ed Hui, Brian Holland, Cynthia Edwalds
- The Society of Actuaries: Korrel Crawford, Erika Schulty, Pete Miller

EVOLVING THE FRAMEWORK FOR MORTALITY IMPROVEMENT

THE JOYS AND WOES OF INSURANCE MI

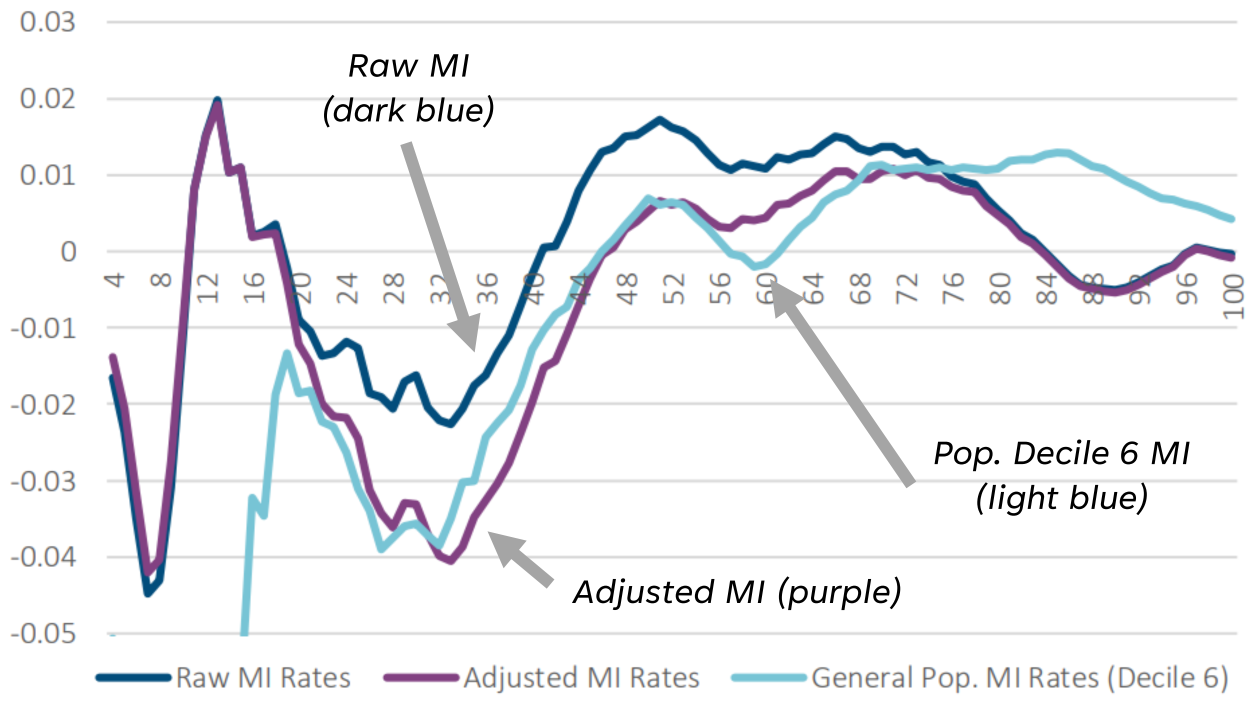
- Insurance MI challenges: lack of long, clean, complete mortality record + credibility
- Application of Framework: elastic net models used to control for exogenous effects
- Framework Complements Traditional Analysis: model identified "nuisance" movements in face amount, preferred class, and insurance plan, which were then controlled for



https://www.soa.org/resources/research-reports/2024/ind-life-mort-tools/

KEY FINDINGS

- Adjusting for externalities (such as preferred risk class) has profound impact
- Traditional Analysis: Female MI over study period more closely resembles population decile 6, as shown to right
- Elastic net: Overall, residual MI after applying elastic net model adjustments reduced top level MI by approximately 40% (not in graph)



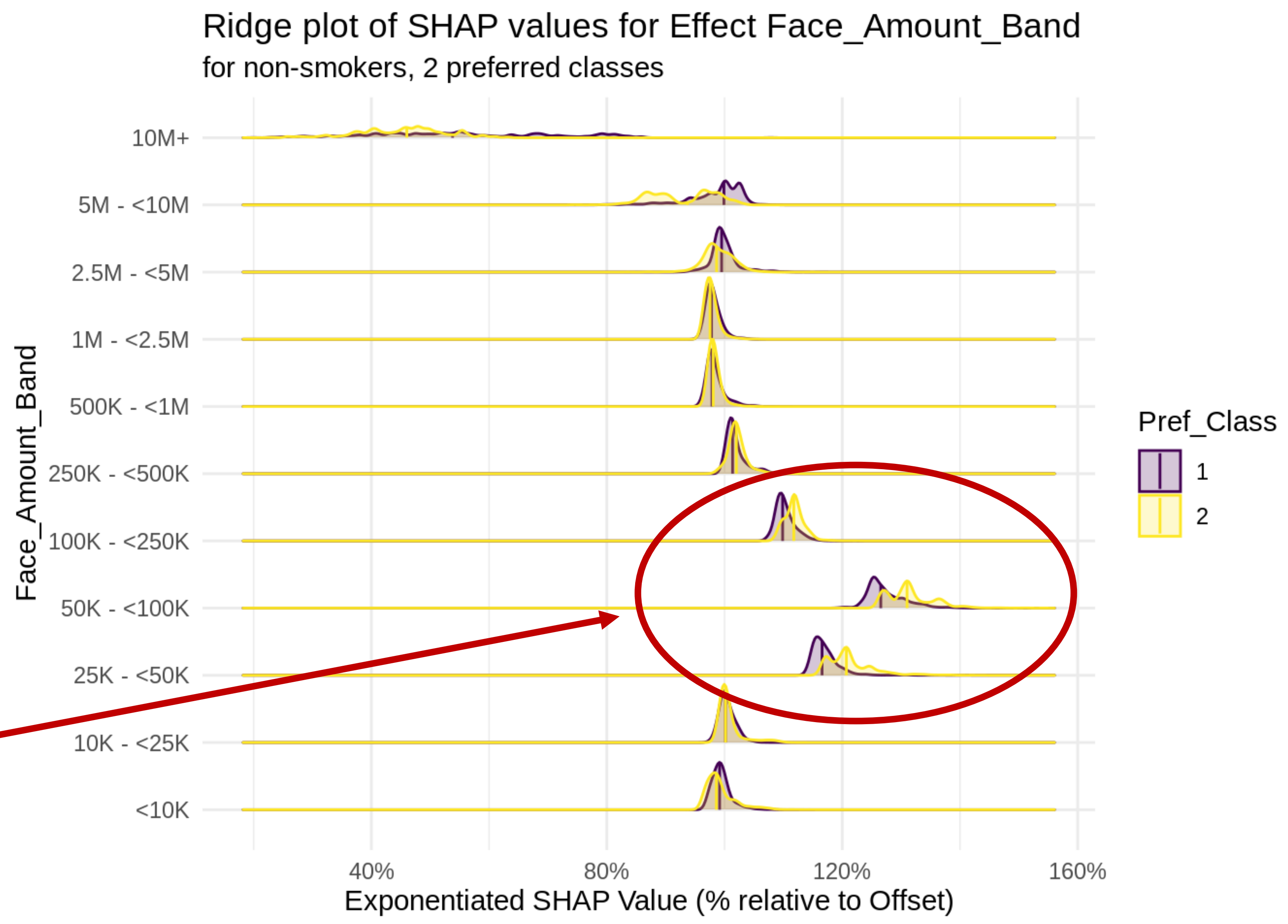
Female Mortality Improvement by Attained Age from Traditional Analysis, "Adjusted MI Rates" are those found after adjusting for preferred risk class information

MORE INFO ON THE SUPPORTING MODEL

1. Use GBDT portion to understand important drivers
   a) Term: underwriting, face amount, duration
   b) Perm: face amount, age, duration
2. Use elastic net GLM for final model
   a) Break the data into five subsets: Perm Unismoke, Perm smoker-distinct, Post-Level Term, Within-Level-Term Under $100,000, Within-Level-Term $100,000+
   b) 1496 parameters (after partitioning)
   c) All interpretable like a GLM, all credible
3. Observation year not included, *but…*
   a) Modeling MI is a challenge (lines, splines, Lee-Carter, etc.)
   b) The goal was to support traditional analysis approach
   c) Leftover trend is less variable across subsets

EVOLVING ILLUSTRATIONS

- Ridge plots illuminate more detail
- Easier to see than boxplots
- Helps detect multiple modes and other differences
- Example: not only are face amount effects different for 25K-250K, but they differ by risk class too



Within-level-Term, face amount SHAP by 2-class preferred

USING ANALYTICS TO ASSIST WITH DATA VALIDATION

THE PROBLEM

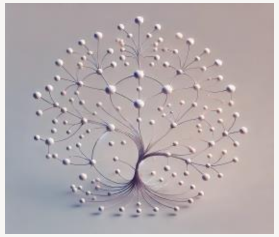
The Setting

- NAIC is the statistical agent for 2018 and later
- For 2018 and 2019, we have over 16 million rows of data, covering 1.2 million death claims
- For years 2011-2017, MIB provided 33 million rows of data covering almost 4 million death claims

The Problem

- Huge datasets: beggars human abilities
- Overpowered: statistical tests will declare even tiny differences significant
- Needle in a haystack: high level may look fine, while obscure corners may differ

TACKLING THE CHALLENGE



Avoiding CSVs: converting data to Arrow Hive	Working in R and Python	GBDT to uncover mortality differences	Vine copulas to uncover exposure differences

DIFFERENCES IN MORTALITY

High Level Patterns

- Differences almost entirely due to improvements in data quality
- Better UW mappings, improvements for PUA handling
- Biggest changes tended to be in fringes

Highlighted Shifts

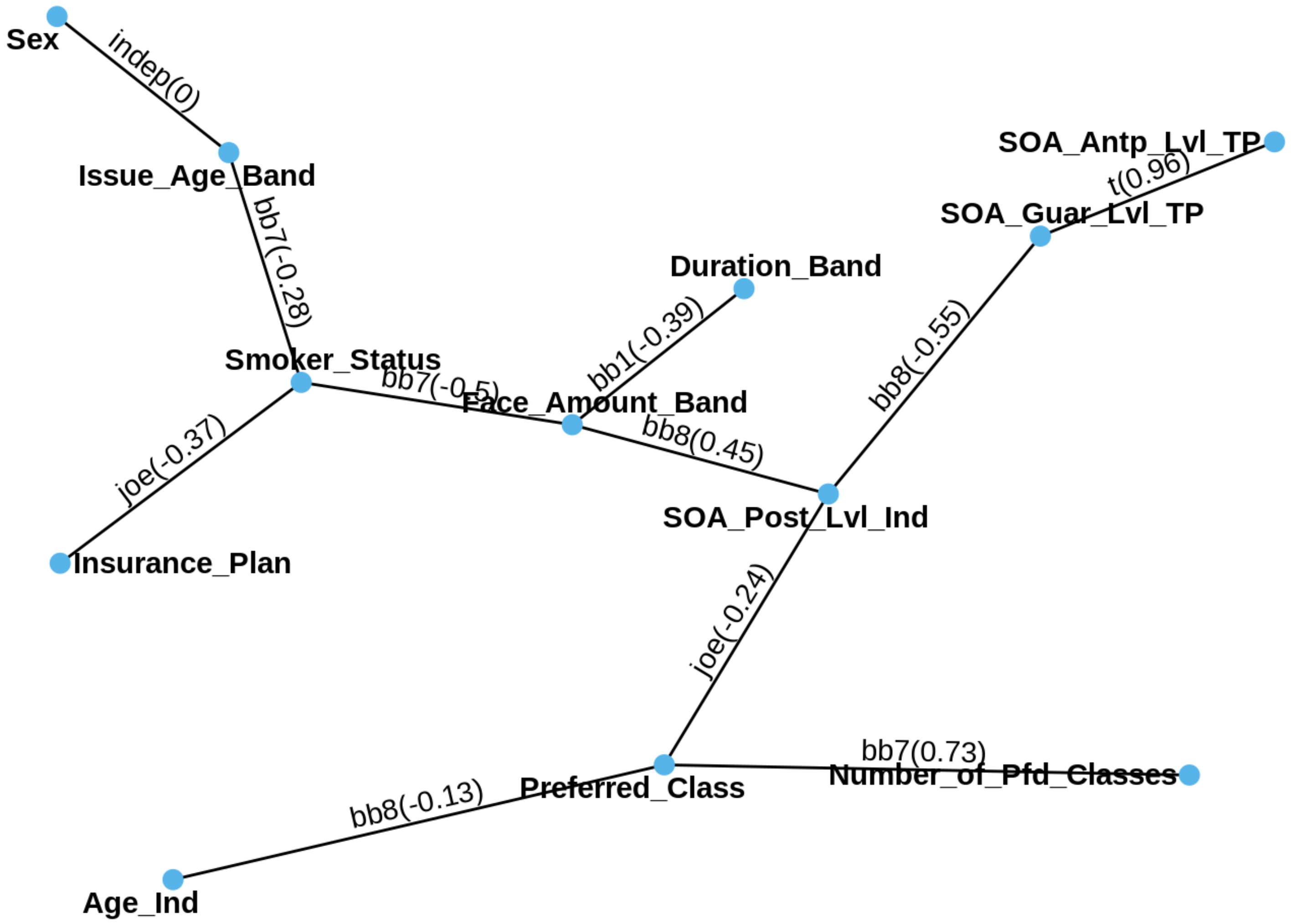
- Perm duration 1 and 2 (↓)
- UL <10K (↑)
- NS 2-class (↓)
- NS 4-class (narrowed spread)
- 25-year term (↓)
- Term Under $100K (↓)

EXPOSURE MODELS WITH VINE COPULAS

- *Every* (cumulative) probability distribution can be rewritten as a (copula) function of its univariate (cumulative) marginal distributions. (Sklar's Theorem)
$$F_{X_1,\ldots,X_n}(X_1,\ldots,X_n) = C\left(F_{X_1}(X_1),\ldots,F_{X_n}(X_n)\right)$$
- Every probability distribution can be decomposed into bivariate copulas in a particular way.
- Turning this around, one can build up probability distributions using bivariate copulas to model distributions (*after making some assumptions).
- Algorithm for vine copulas
    - Find graph of strongest pairwise dependencies between variables
    - Fit copula for the edges of that graph (i.e., for strongest dependencies)
    - Copula on edge becomes new "variable"
    - Repeat until no more pairs to model

DIFFERENCES IN EXPOSURES

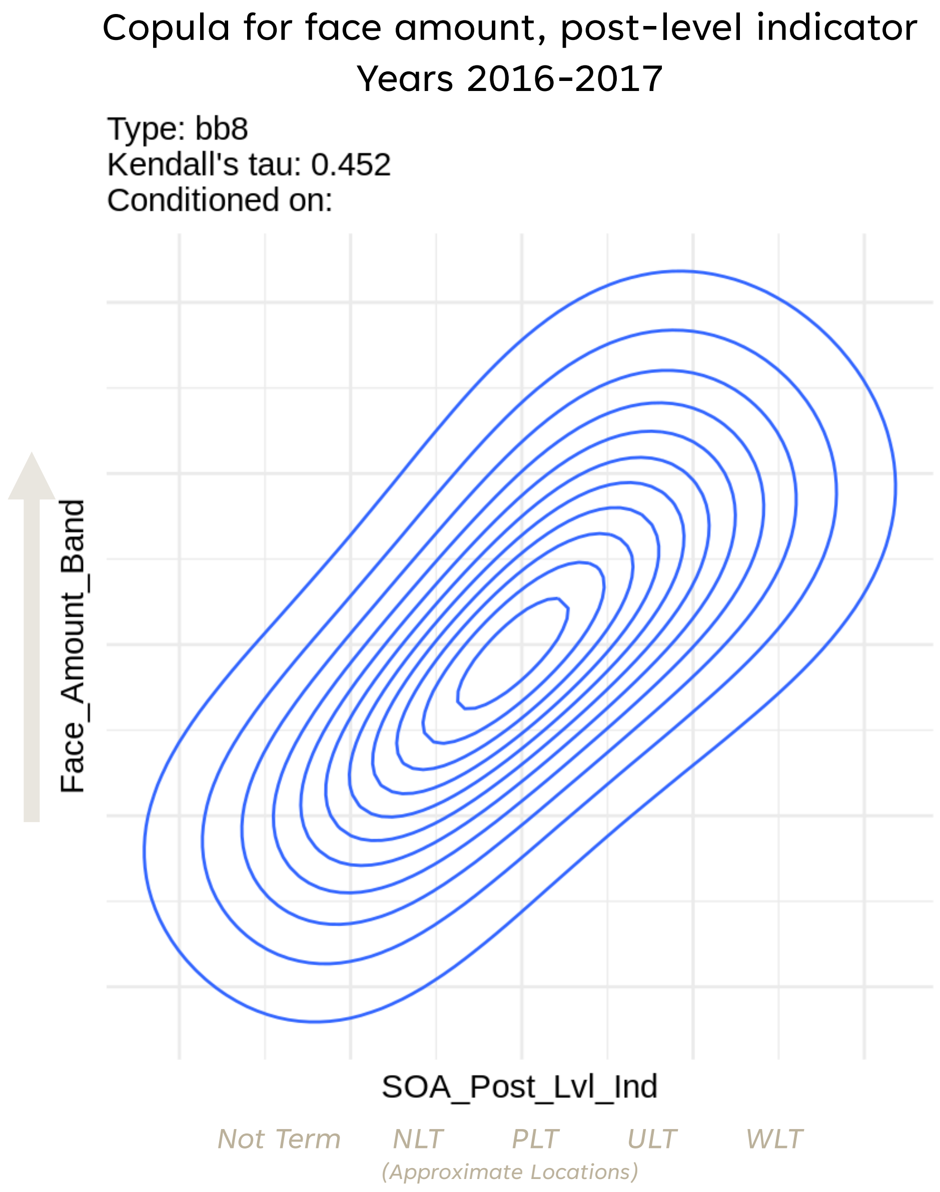
- Some drift in univariate marginal distributions were noted, especially by plan
- By-count copulas were almost identical across years, save for emergent interdependency among sex, face amount band, plan, and UW
- Two-way interactions were usually enough



Dependency Graph of Two-way, Unconditional Interactions for Policies Exposed, years 2016-2017

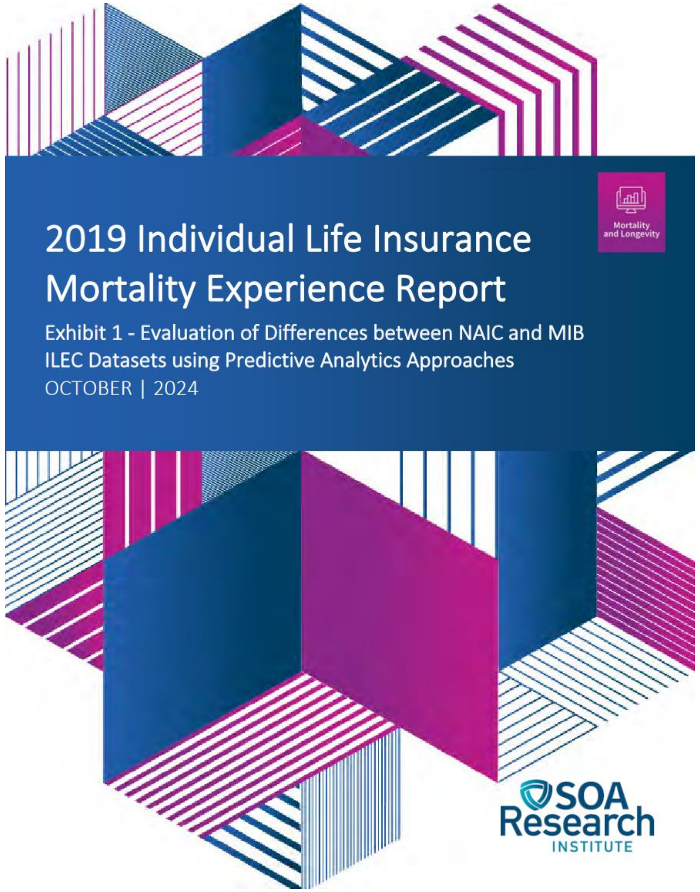
ZOOMING IN: FACE AMOUNT AND POST-LEVEL INDICATOR

- Post-level Indicator overlaps with Insurance Plan
- Higher face amounts associated with within-level-term, unknown level term (i.e., ART on level term)
- Lower face amounts associated with not level-term, “not term”, and post-level term



SEE FOR YOURSELF

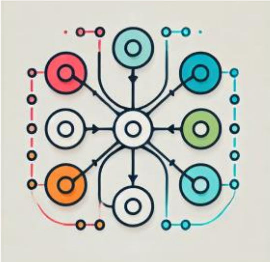
- Extensive details on mortality and exposure differences in Exhibit 1 and the Vine Copula Models supplement (with code!)

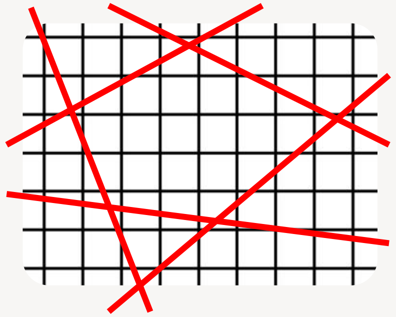


https://www.soa.org/4a8d7e/globalassets/assets/files/resources/research-report/2024/ilec-mort-exhibit1.pdf

CAN NEURAL NETS WORK FOR MORTALITY?

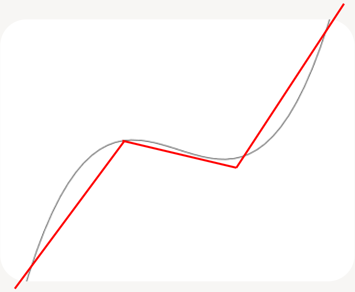
NEURAL NETS





Term
UL
10YT
…

?



Feed forward neural nets of rectified linear units: dynamically partition input space, with a linear model in each partition

ILEC data is a regularly spaced grid, so hard to partition

Neural nets have trouble with category-heavy mortality data

NNs have trouble with non-linear relationships

Examples of using NNs on mortality data sometimes use wrong loss function, *even in published papers*

MAKING IT WORK

Challenges and Solutions

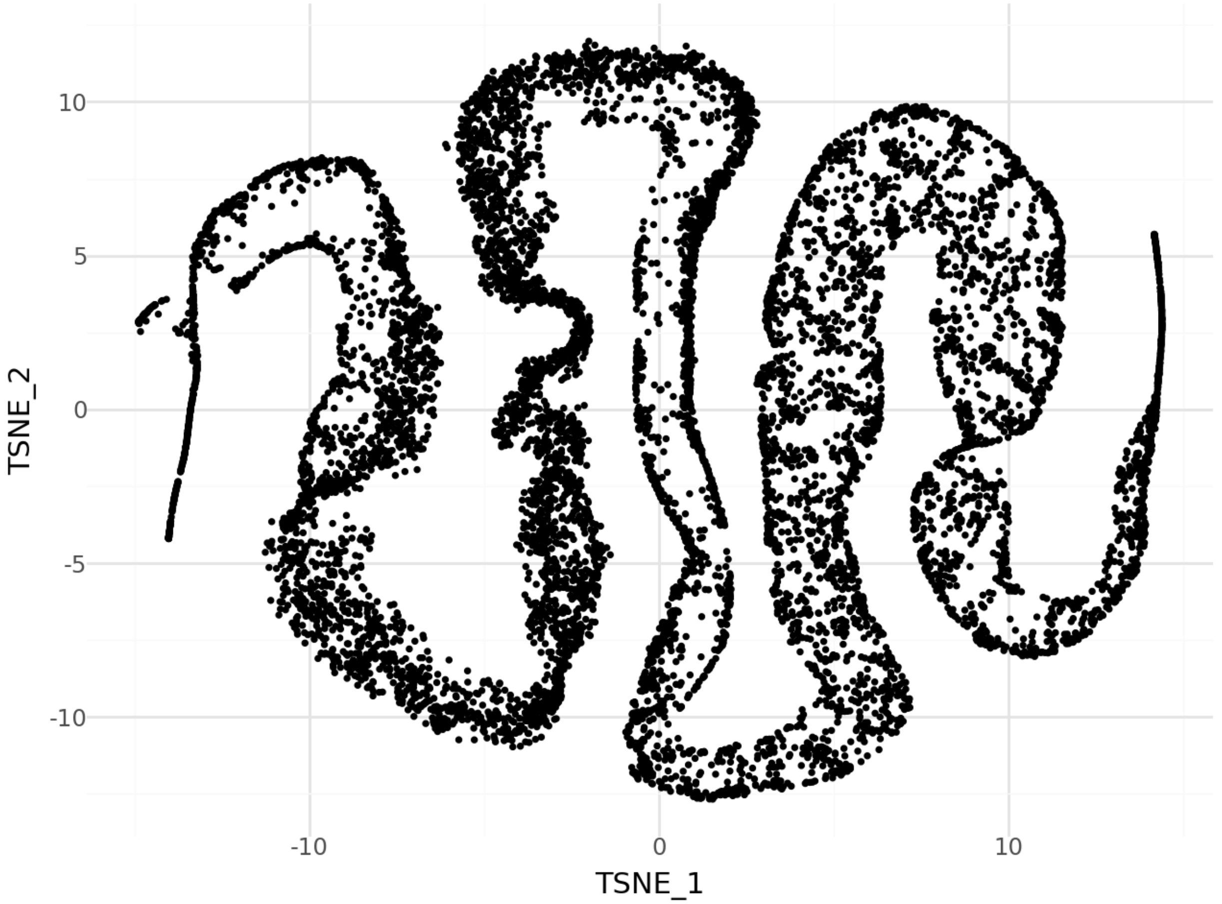
- Use the right loss function
- Experiment with categorical embeddings
- Transform “continuous” variables to lend the NNs a hand
- Secure enough computational resources

For This Project

- ~ 4 million input rows of data
- ~ 10,000 distinct categorical combinations
- Three NN layers
- About 15-20 minutes and ~ 20 GB VRAM for 5000 iterations on RTX 4090

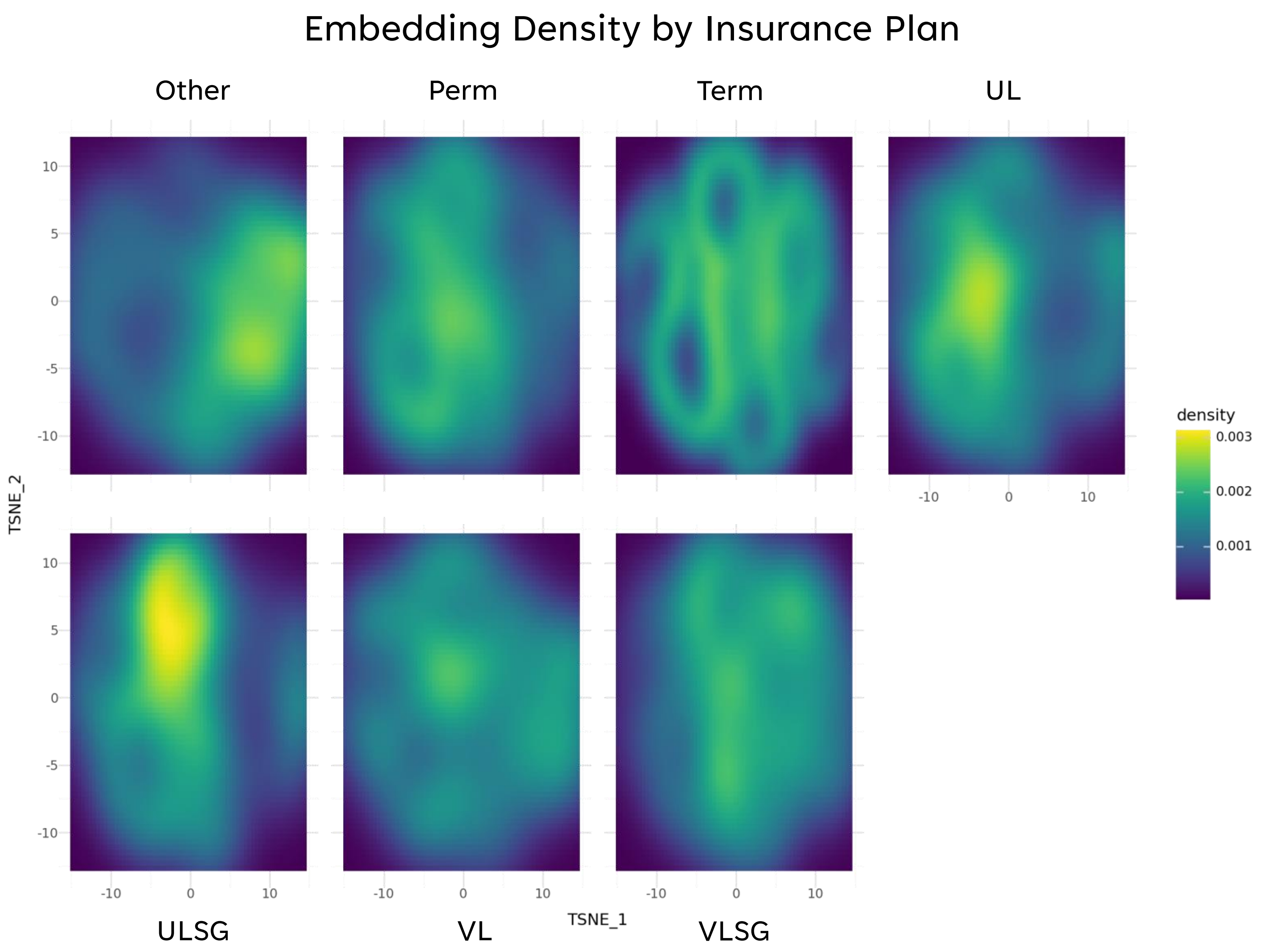
EARLY RESULTS: EMBEDDINGS

- Embedding maps each combination of nine categorical predictors into a 9-D free vector
- NN adapts embeddings to become coefficients for spline basis
- Collection of fitted 9-D vectors is mapped to 2-D via tSNE



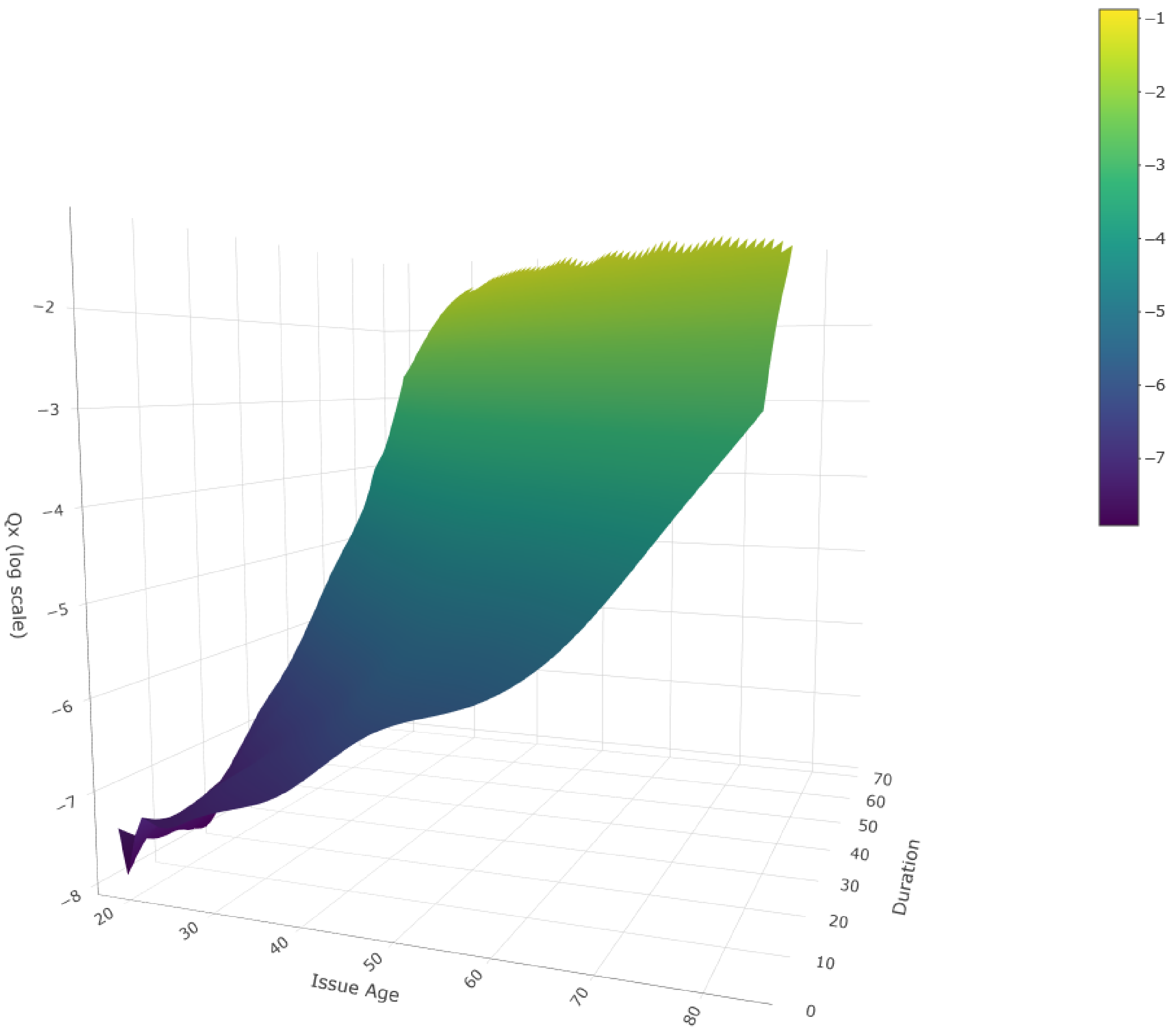
EARLY RESULTS: EMBEDDINGS

- There is structure in the embeddings
- Term and Other tend to be on the right
- Perm, UL, ULSG tend to be on the left



EARLY RESULTS: GENERATED TABLES

- Generated tables may be believable, but may or may not be credible
- Example at right:
  - 197 claims in cell
  - 131 of 680 age/duration combos have claims
- Why so nice? Information from other cells informs the estimate



M, NS/2/2, Term, 100K-249K, ANB, 10 YT Ant, WLT

LOOKING TO THE FUTURE

ILEC provides useful tools and methods for actuaries to use in their work

The door is now open for neural net applications and experiments on insurance mortality data (e.g., CNN, RNN, LSTM, etc.)

Analytics and machine learning will be integrated into more and more of ILEC's work