Mortality and Longevity

A Predictive Analytics Framework

A Reusable Workflow For Experience Analysis

NOVEMBER | 2024

SOA Research INSTITUTE

A Predictive Analytics Framework

A Reusable Workflow for Experience Analysis

AUTHOR Philip L. Adams, FSA, MAAA
Cynthia Edwalds, FSA, ACAS, MAAA
Brian Holland, FSA, MAAA
Michael Niemerg, FSA, MAAA
Haofeng Yu, FSA, MAAA

SPONSORS Mortality and Longevity Strategic Research Program Steering Committee

CONTENTS

A Predictive Analytics Framework

A Reusable Workflow for Experience Analysis

Executive Summary

Predictive modeling is a powerful tool that leverages statistical techniques to forecast outcomes. It is widely used across various industries, from healthcare to finance, to make informed decisions based on historical data. This tutorial aims to introduce a framework for developing predictive models in the context of actuarial experience studies. To ground this framework within the context of real actuarial problems, we will also specifically look to understand and model the differences in mortality by product (whole life, term, etc.) With the modeling approach, we can see that.

- The relative spread of preferred mortality differs by product.
  - For two-class preferred systems, the residual standard mortality is much higher than preferred for term than for other products.
  - The spread for UL/VL/ULSG/VLSG for four-class preferred systems is much wider than for other products.
- There are divergences in the spread of face amount factors for xL, Perm, and Term, with xL narrowing relative to Term.
- The issue age slope appears to be steeper for Term than Perm and xL under age 65. However, differences emerge above issue age 65, with the slope for Perm steepening relative to xL.
- There are differences among the products in durations 1 and 2.
- Since issue years 1990-1999, there has been a small but steady increase in relative mortality for xL versus Term, with xL now approaching Term.

Section 1: Data

For what follows, we used a filtered and summarized subset of the Society of Actuaries Research Institute's Individual Life Experience Committee (ILEC) mortality data. Columns included in the extract were:

- Number of Preferred Classes
- Preferred Class
- Smoker Status
- Face Amount Band
- Observation_Year
- Duration
- Issue Age
- Insurance Plan
- Anticipated Level Term Period
- Issue Year
- Sex
- Death Count
- Death Claim Amount
- Tabular Expected Mortality by Count - 2015VBT
- Tabular Expected Mortality by Amount - 2015VBT

The data were filtered as:

- Issue ages 18 and greater
- Durations 25 and less
- Experience years 2013-2017

and then grouped or combined as:

- Underwriting: concatenation of smoker status, number of preferred classes, and preferred class, in that order
- Duration: 1, 2, 3, 4-5, 6-15, 16-25
- Issue Age: 18-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85-99
- Issue Years: 1900-1989, 1990-1999, 2000-2009, 2010+
- Insurance Plan: UL, ULSG, VL, VLSG collapsed into category "xL"
- Face Amount Band: face amounts under $50,000 grouped into a single category, face amounts $1 million and greater grouped into a single category

The intent of this heavy grouping and summarization was to enable running this document with modest computing resources. The source data can be replaced with a similarly constructed dataset with more finely grouped variables.

The code to generate these files can be found in the datafiles subfolder. It relies on an unpublished version of the ILEC dataset, which has been restructured using the Arrow framework into a collection of Parquet files. A knowledgeable reader should be able to adapt the code to whatever environment in which they keep their own copy of the ILEC dataset.

Section 2: Machine Learning in Mortality Studies

Experience studies are a primary tool that actuaries use to quantify and understand historical experience. It is a natural next step to apply statistical techniques to experience studies to discover new and relevant insights. There are multiple advantages to this approach:

- Allows the actuary to avoid cumbersome and potentially misleading univariate analysis
- Allows the actuary to appropriately consider credibility and unlock all the credibility inherent in the data
- Makes it easier to discover and appropriately adjust for variable interactions
- Enables the actuary to statistically control for the different sources of variation in any given cell of a mortality study.

2.1 PROBLEM STATEMENT

One question of interest to actuaries is why different products exhibit different mortality outcomes. Even though they can be difficult to separately identify and quantify, it is known that underwriting, target market, policyholder behavior, and socioeconomic factors, among others, have a direct bearing on mortality outcomes. With a statistical or machine learning model, we have a possible solution to account for the impact of these variables. For this project, the key question we are trying to answer is how mortality varies by product in the Individual Life Experience Committee dataset. In the simplified dataset that is used herein, the product categories are Term, Perm, UL/VL, and Other. To understand the differences in mortality by product, we will construct machine learning models to predict the mortality outcomes and analyze the results for relevant insights.

2.2 METHODOLOGY

The framework will guide the process of code setup, model creation, preprocessing, and validation. It will also address common challenges often encountered such as: incorporating nonlinear relationships, determining interactions, dealing with underfitting and overfitting (bias-variance trade-off), and model interpretability. The goal of this project is to provide useful techniques, code, and ideas to actuaries to guide future analysis of mortality studies.

There are several common key steps in any modeling process: data preprocessing, data exploration, model selection, model validation, and model interpretation. Much more can be written on these topics than we have the space to explore, and we aim to address the key considerations as they pertain to experience studies.

2.3 MODELING APPROACHES

When applying statistics and machine learning to experience studies, there are multiple different modeling approaches one might take. We will focus our attention on the most common approaches used: generalized linear models (GLMs), generalized linear models with penalization (also known as elastic net GLMs), and gradient boosting machines (GBMs or GBDTs). Many other approaches or variations on these approaches are also reasonable.

2.3.1 GENERALIZED LINEAR MODELS (GLM)

Generalized linear models have the most history of the methods that we will examine and, in some sense, are the simplest. One of the benefits of GLMs is that they allow statistical hypothesis testing. For instance, individual model coefficients can be statistically tested, and various statistical tests can be performed to validate results and compare models. The results of GLMs are also relatively simple to interpret. However, GLMs have a few disadvantages: due to their relative simplicity, they have lower predictive power than other methods. To get the best performance out of a GLM, additional effort is needed to capture nonlinear relationships and interactions. Ultimately, this can make them more time-intensive than other methodologies.

GLMs can be extended into regularized GLMs, such as LASSO or Ridge, which modifies the objective used to fit the model. This regularization term offers several advantages, disadvantages, and changes to the modeling process. First, the addition of penalization makes confidence intervals and hypothesis testing infeasible. Instead of using hypothesis testing on coefficients and likelihood ratio tests to evaluate relative fitness of models, we apply a machine learning paradigm by optimizing our model using cross-validation. Fortunately, a regularized GLM still maintains the nice interpretability of a linear model, and it can increase the overall predictive accuracy of the model. Additionally, by using a LASSO penalty, it can perform automatic feature selection.

2.3.2 GRADIENT-BOOSTED DECISION TREES (GBDT)

Gradient boosting decision trees are an ensemble of decision trees generated in a stage-wise fashion. Each decision tree is recursively trained on the residuals of the previous tree. The first tree is a decision tree on the outcome, the second the residuals on that, and so on. In this way, the model is continually refocusing on where its predictions are weakest. Popular frameworks for gradient boosting decision trees include LightGBM, CatBoost, and XGBoost. This model is one of the most effective methods for classification and regression for tabular data.

Gradient boosting machines (applied here with LightGBM) have become the go-to approach in many tabular machine learning tasks due to their very high accuracy, ease of use, and ability to seamlessly discover important interactions. However, they can also be the most complex to interpret. To aid in interpretation, we will discuss the use of SHAP values, which is a popular method of interpretation.

2.4 MODEL EXPLANATION

2.4.1 ORDERED LORENZ PLOT AND GINI

An ordered Lorenz curve and the associated Gini coefficient measure the ability of a model to stratify risk. An ordered Lorenz curve is created using the model prediction as an index. Using this index, we graph the cumulative percentage of claims versus the cumulative percentage of exposure. The more bowed this line is, the better the model is able to predict the outcome. The Gini Index measures the difference between this line and perfect equality. The more your model is able to predict risk, the more unequal the distribution of claims is between the model prediction and, thus, the larger the Gini coefficient.

2.4.2 LIFT PLOT

There are several different varieties of lift plots used in connection with machine learning. These plots are used to help visually understand the risk stratification and accuracy of a model. As presented here, lift plots sort the model predictions into deciles based upon the predicted value. For each decile, the model's average prediction for that cell is graphed versus the value seen empirically in the data. The more these two values are in agreement, the better the model is performing.

2.4.3 SHAP

SHAP values are a method of model interpretation in machine learning and originally come from Shapley values in economics. SHAP values measure the impact each feature has on the prediction for a particular instance. This numeric score indicates how much each feature contributed to the prediction in terms of sign and magnitude.

2.4.4 FEATURE IMPORTANCE

Feature importance is a global measure of how much a variable contributes to the predictions of a target variable within a model. This can be helpful in interpreting a model to understand the key drivers in aggregate. However, unlike SHAP values, feature importance does not help you interpret individual predictions. Feature importance is usually presented in terms of percent contribution. When done so, a feature importance of 20% for a feature would imply that 20% of the overall reduction in prediction error is attributable to that particular feature. There are multiple ways of measuring feature importance. One of the simplest and most intuitive is permutation feature importance. Using this method, you scramble a particular feature so that it is no longer useful and measure the percent difference in model performance before and after this change. The change in error would be the importance.

The reader should be cautioned that a low relative importance does not imply lack of significance or of predictive value. For example, gender is a well-known predictor of mortality. The variation explainable by other factors of the data can greatly exceed the variation arising from gender, and interactions with other variables like age can further rob gender of importance attributed to it. The effect then is to push gender down the feature importance list.

2.4.5 GOODNESS-OF-FIT

No matter how well a model may behave on measures of feature importance, lift, Lorenz and Gini indices, mean square error, deviance, and so on, it is nonetheless important to check goodness-of-fit. Goodness-of-fit checks allow us to see how well a model reproduces the phenomena of interest. For our purposes, this is the same as checking ratios of actual claims to model predicted claims. In each model section, there are univariate and bivariate goodness-of-fit tables. Ideally, we should see 100% for all entries. For the GLM model and the univariate goodness-of-fit checks, we will see this throughout the tables of goodness-of-fit, as a non-penalized GLM will reproduce the margins for any included categorical variable or interaction of categorical variables. For space reasons, we omit a test for ratios significantly different from 100%. However, qualitatively, ratios far from 100%, perhaps +/- 5% or +/- 10%, should be deemed as evidence of poor fit for that cell.

Section 3: Framework Preparation

Before getting to the core data analysis task, we need to first prepare the R environment by configuring display and model options, loading necessary libraries. Then, we load the data and prep for running data analysis and modeling.

Early in the framework, we set parameters which control the subsequent operation of the workflow. In the report, we document these in an appendix.

When we read the data, we convert specified columns to categorical factors to ensure proper data handling and adjust the labels for the 'face_amount_band' factor to avoid issues in model outputs. The dataset is split into training and testing sets based on the observation year, with the year 2017 data used for validation.

Section 4: Models

4.1 GLM

Below is an analysis using a main-effects GLM to better understand the data. We integrate two modeling approaches:

- Standard GLM analysis with model calibration (but no model building), including presentation of coefficients and residuals analysis, and
- An approach to explore the interactions of the main effects model along selected dimensions of the data, checking the average main effects for those subsets, weighted according to the offset used in the analysis.

4.1.1 MODEL SUMMARY

Model Summary

Below is the table of coefficients for the fitted GLM. Each entry is a coefficient in the table for the level of the indicated variable. The estimate and standard errors are on the scale of the linear predictor. For a Poisson model with log link, this means they are on the log scale.

Coefficient	Estimate	Standard Error	z value	Pr(>|z|)	
(Intercept)	0.978	0.091	10.743	0.0000	***
uwN/2/1	-0.152	0.011	-13.931	0.0000	***
uwN/2/2	0.248	0.010	23.945	0.0000	***
uwN/3/1	-0.324	0.014	-23.614	0.0000	***
uwN/3/2	-0.180	0.012	-14.802	0.0000	***
uwN/3/3	0.152	0.011	14.077	0.0000	***
uwN/4/1	-0.330	0.014	-23.555	0.0000	***
uwN/4/2	-0.166	0.015	-11.060	0.0000	***
uwN/4/3	0.011	0.017	0.630	0.5287	
uwN/4/4	0.208	0.016	12.618	0.0000	***
uwS/1/1	0.068	0.014	4.778	0.0000	***
uwS/2/1	-0.179	0.021	-8.637	0.0000	***
uwS/2/2	0.112	0.022	5.178	0.0000	***
uwU/1/1	0.261	0.038	6.794	0.0000	***
face_amount_band04 - 50,000 - 99,999	-0.113	0.018	-6.410	0.0000	***
face_amount_band05 - 100,000 - 249,999	-0.236	0.015	-15.413	0.0000	***
face_amount_band06 - 250,000 - 499,999	-0.299	0.016	-18.893	0.0000	***
face_amount_band07 - 500,000 - 999,999	-0.322	0.016	-20.258	0.0000	***
face_amount_band08 - 1,000,000+	-0.349	0.015	-23.004	0.0000	***
dur_band102	0.010	0.030	0.341	0.7334	
dur_band103	0.000	0.029	0.006	0.9951	
dur_band104-05	-0.044	0.026	-1.686	0.0917	.
dur_band106-15	-0.115	0.028	-4.084	0.0000	***
dur_band116-25	-0.096	0.031	-3.088	0.0020	**
ia_band125-34	-0.062	0.034	-1.813	0.0699	.
ia_band135-44	-0.045	0.033	-1.347	0.1779	
ia_band145-54	-0.075	0.033	-2.258	0.0239	*
ia_band155-64	-0.100	0.033	-3.024	0.0025	**
ia_band165-74	-0.082	0.033	-2.440	0.0147	*
ia_band175-84	-0.173	0.034	-5.100	0.0000	***
ia_band185-99	-0.304	0.040	-7.581	0.0000	***
genderM	0.011	0.006	1.790	0.0735	.
insurance_planPerm	-0.134	0.062	-2.159	0.0308	*
insurance_planTerm	-0.275	0.069	-3.977	0.0001	***
insurance_planxL	-0.042	0.061	-0.681	0.4958	

Coefficient	Estimate	Standard Error	z value	Pr(>|z|)	
ltp10 yr	-0.090	0.034	-2.650	0.0080	**
ltp15 yr	-0.097	0.034	-2.826	0.0047	**
ltp20 yr	-0.207	0.033	-6.286	0.0000	***
ltp25 yr	-0.247	0.049	-5.077	0.0000	***
ltp30 yr	-0.161	0.036	-4.502	0.0000	***
ltpNot Level Term	-0.396	0.045	-8.750	0.0000	***
ltpUnknown	-0.207	0.036	-5.804	0.0000	***
iy_band11990-1999	-0.073	0.023	-3.209	0.0013	**
iy_band12000-2009	-0.148	0.026	-5.706	0.0000	***
iy_band12010+	-0.206	0.030	-6.857	0.0000	***

*Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05*

(Dispersion parameter for quasipoisson family taken to be 700118)
Null deviance: 6.172e+10 on 249979 degrees of freedom
Residual deviance: 5.662e+10 on 249935 degrees of freedom

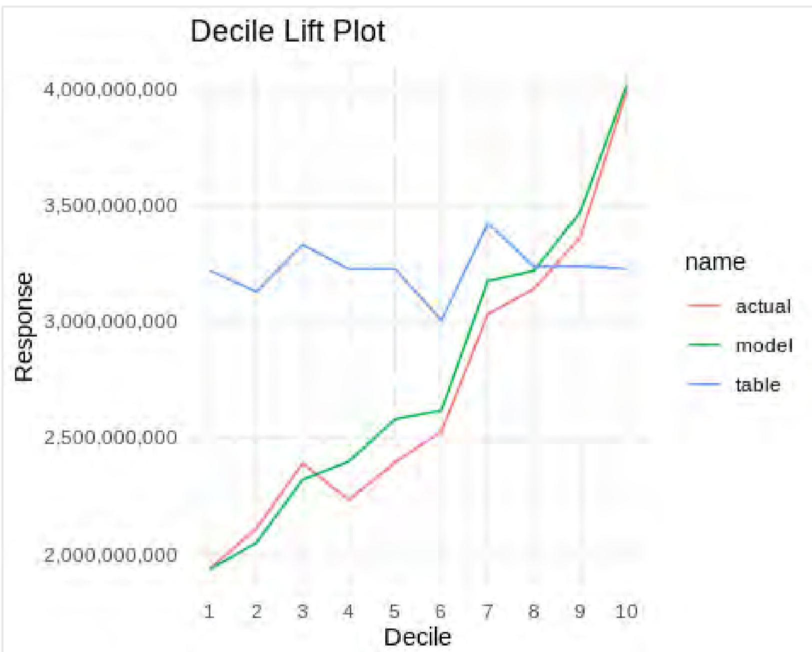
ANOVA

The ANOVA table displays the analysis of deviance for the GLM. For each variable, we see the proportion of deviance explained by that variable and its associated degrees of freedom.

feature	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			249,979	61,716,566,139	
uw	13	3,858,650,451	249,966	57,857,915,688	< 0.1%
face_amount_band	5	701,275,400	249,961	57,156,640,288	< 0.1%
dur_band1	5	118,698,616	249,956	57,037,941,672	< 0.1%
ia_band1	7	162,999,155	249,949	56,874,942,517	< 0.1%
gender	1	3,395,713	249,948	56,871,546,804	2.76%
insurance_plan	3	57,762,926	249,945	56,813,783,878	< 0.1%
ltp	7	152,501,012	249,938	56,661,282,865	< 0.1%
iy_band1	3	41,049,822	249,935	56,620,233,043	< 0.1%

Lift

The lift plot compares the GLM against the underlying mortality table.



Lorenz Plot

The Lorenz plot demonstrates a model's ability to stratify predictions against a null baseline.

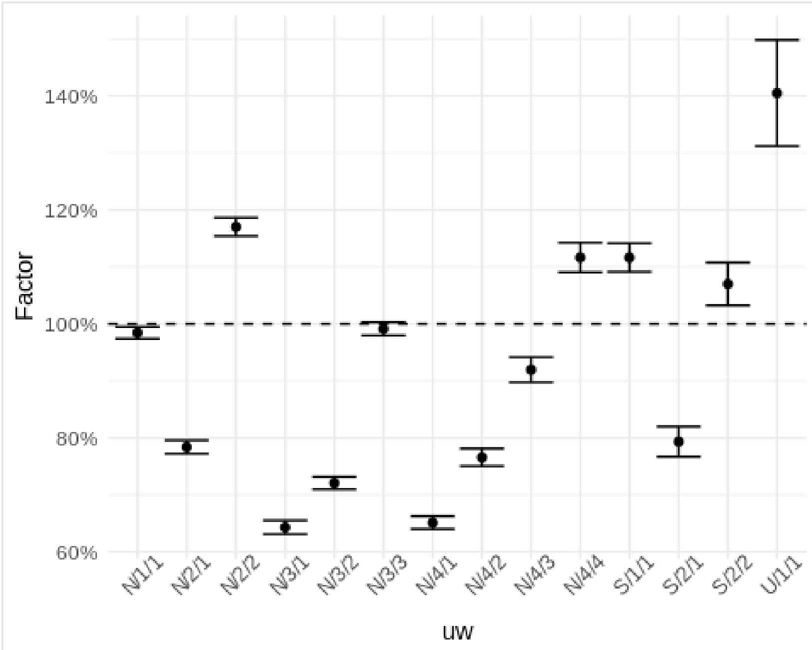
The table of coefficients shows a number of interesting phenomena and, perhaps, some surprises:

- Gender is not significant. Since we are using an offset of tabular expected rates, the interpretation is that the underlying differentials in the tabular expected rates are adequate for the current data, after adjusting for other factors.
- Underwriting is the most influential factor from the ANOVA perspective.
- Both the most recent issue years (2010+) and the most recent durations show significant mortality factors. Durations 1 and 2 are significantly higher than durations 3+
- While insurance plans other than "Other" are significantly different from 0, a quick glance at the effects plot shows that the UL/VL plans are not significantly different from each other and with Perm, while Term is borderline significantly different from UL/VL.
- Face amount bands $250K and greater have factors not significant from one another.

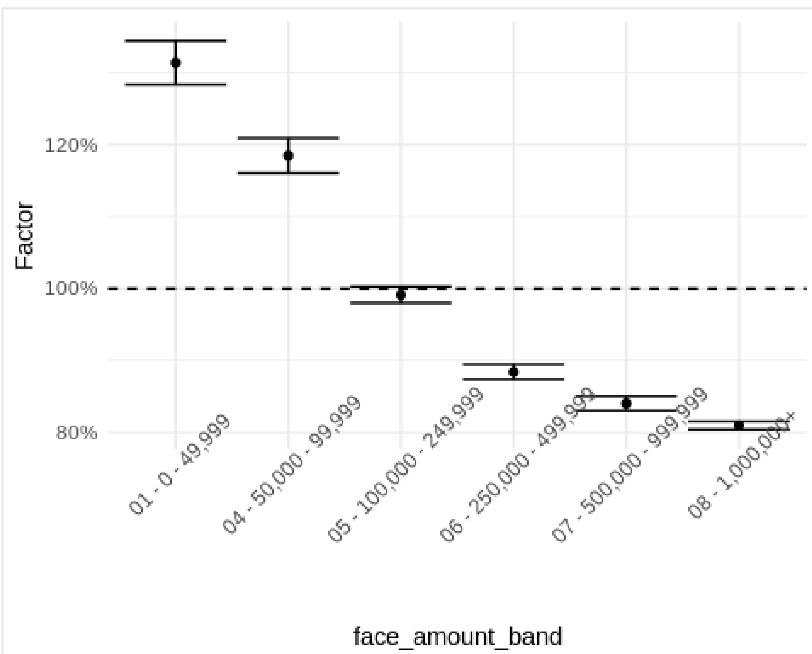
4.1.2 MODEL ILLUSTRATIONS AND GRAPHICS

Effects Plots

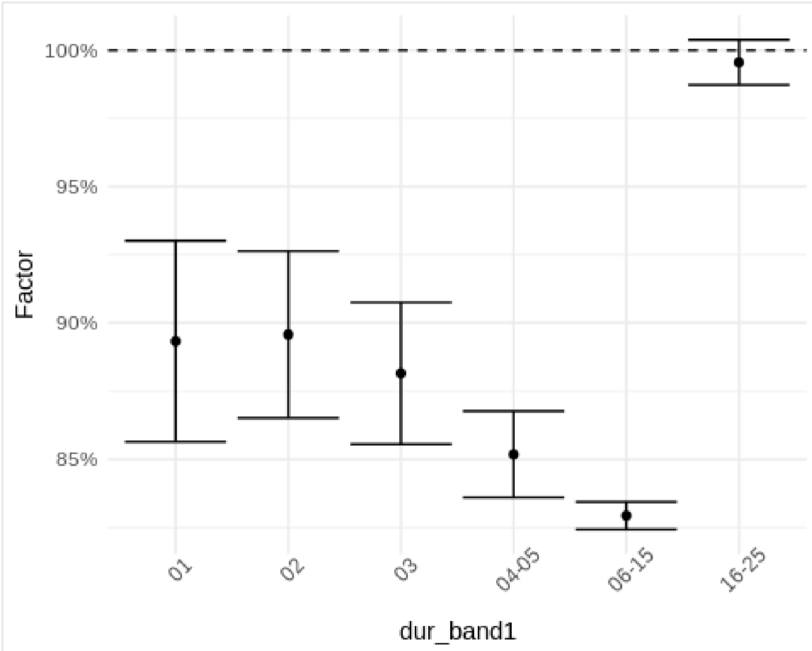
Because the model contains every column, this is equivalent to computing the marginal actual-to-tabular ratios. However, the model also provides standard errors, which is useful for assessing the significance of the marginal ratios.

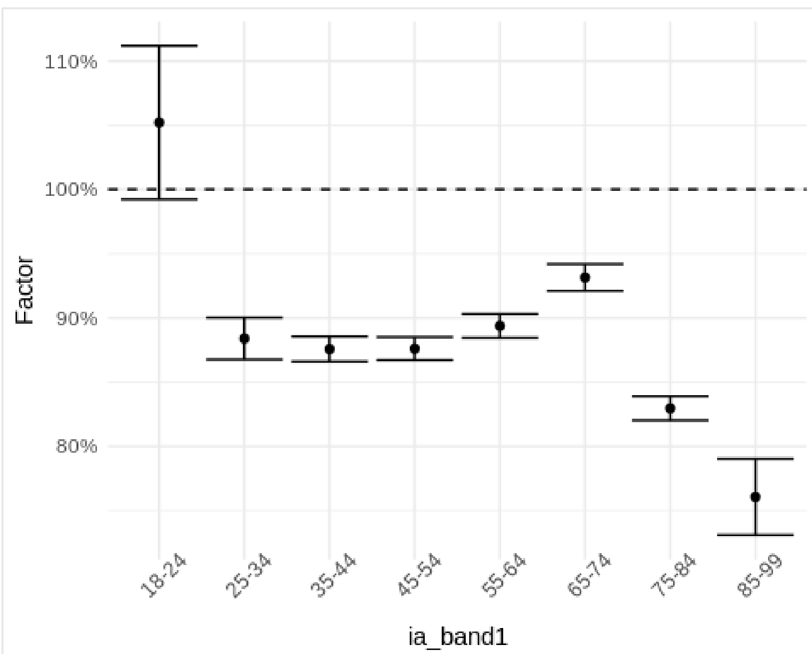


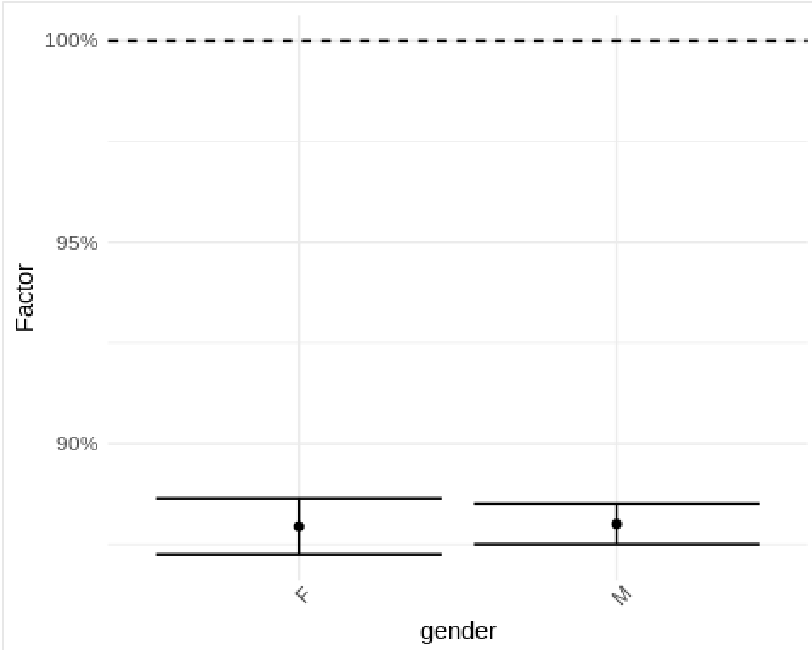
140%
120%
100%
80%
60%
Factor
N/1/1
N/2/1
N/2/2
N/3/1
N/3/2
N/3/3
N/4/1
N/4/2
N/4/3
N/4/4
S/1/1
S/2/1
S/2/2
U/1/1
uw

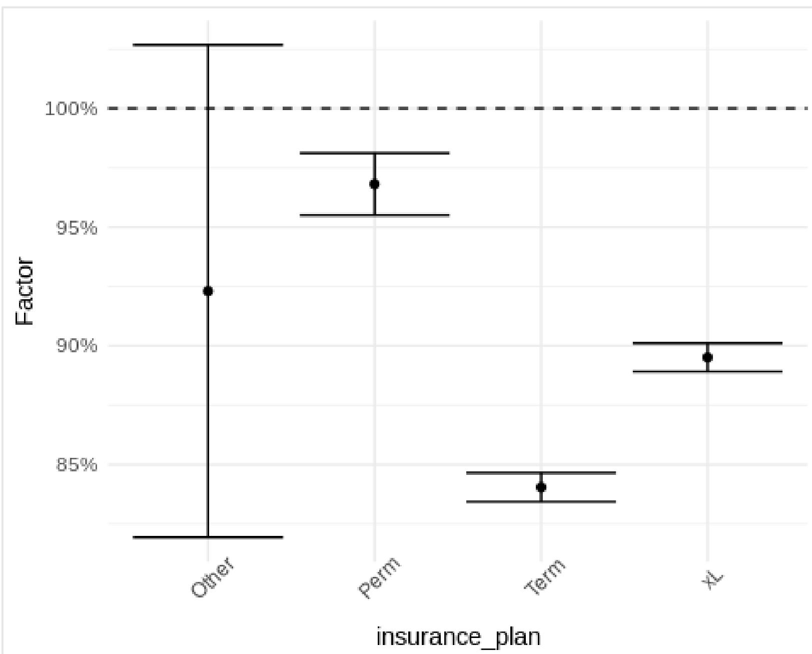


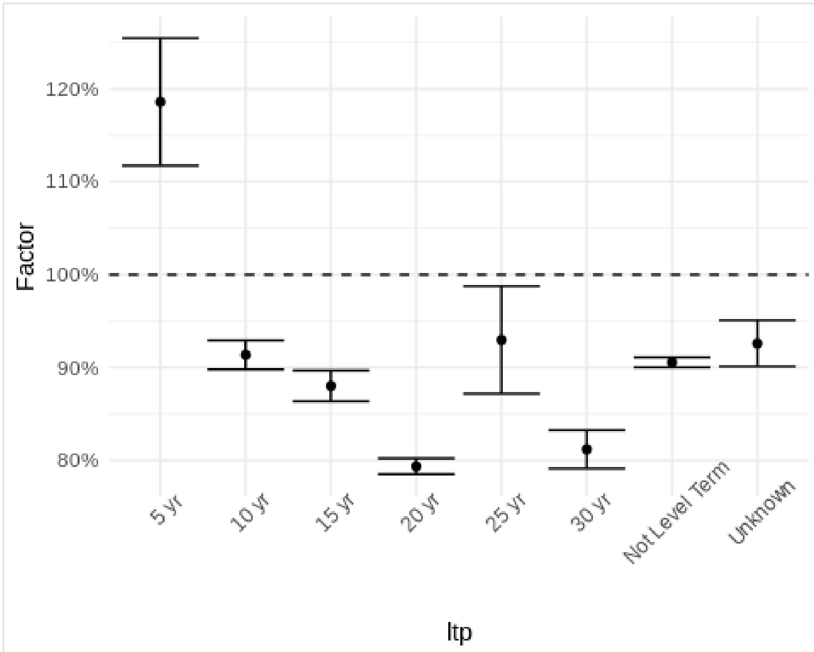
120%
100%
80%
Factor
01 - 0 - 49,999
04 - 50,000 - 99,999
05 - 100,000 - 249,999
06 - 250,000 - 499,999
07 - 500,000 - 999,999
08 - 1,000,000+
face_amount_band

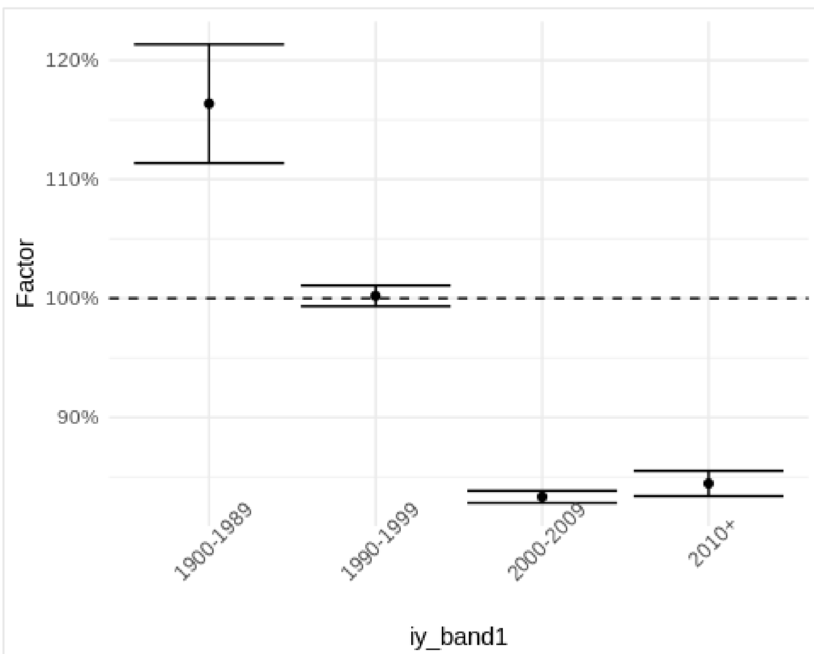

100%
95%
90%
85%
Factor
01
02
03
04-05
06-15
16-25
dur_band1


110%
100%
90%
80%
Factor
18-24
25-34
35-44
45-54
55-64
65-74
75-84
85-99
ia_band1


100%
95%
90%
Factor
F
M
gender


100%
95%
90%
85%
Factor
Other
Perm
Term
xL
insurance_plan





Goodness-of-Fit

Goodness-of-fit tables are provided. Each table provides actual-to-model ratios for single variables and for two-way combinations of variables. A model is qualitatively deemed to perform well if goodness-of-fit ratios are close to 100% in almost all situations. The quantitative assessment using significance testing is omitted here.

See the Excel tables in Exhibits 1 and 2 accompanying this report for additional information.

Subgroup Variability

This section reproduces Brian Holland's publication. For background on the tables generated below, please refer to the publication.

See the Excel tables in Exhibit 3 accompanying this report for more information.

4.2 LIGHTGBM

4.2.1 DATA PREPARATION

First, the data are prepared for LightGBM. LightGBM expects matrices for its inputs. Thereafter, the LightGBM model is trained. Factors are recast as their underlying integer indices.

4.2.2 MODEL FITTING

The LightGBM model is fit to the training subset using a Poisson objective. The model response is the ratio of response variable and response offset, and the weights are the specified offset. Often, this might be “actual claims” as the response and “expected claims” as the offset.

4.2.3 MODEL ILLUSTRATIONS AND GRAPHICS

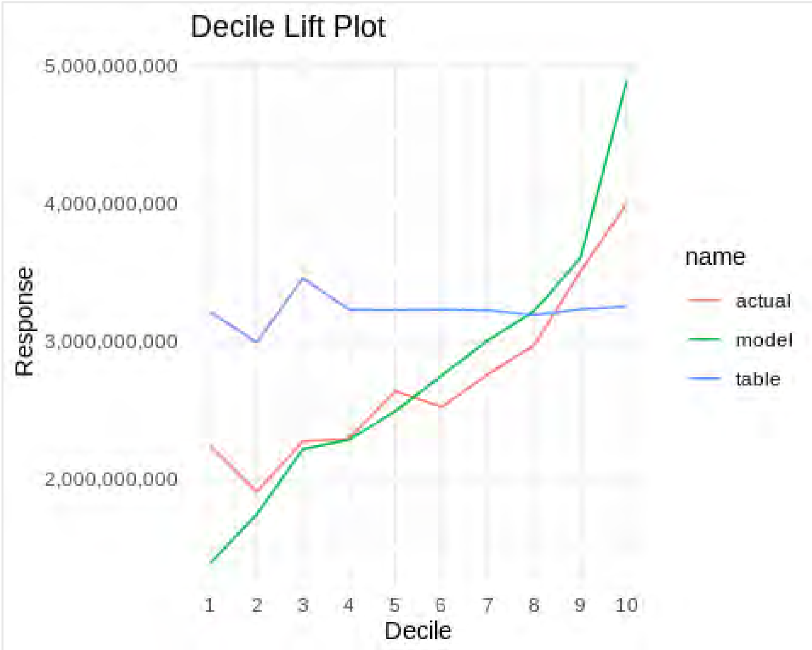
From this, we can plot decile lift and Lorenz curves.

The decile lift plot can be interpreted as a way to visualize the effectiveness of a predictive model. It divides the data into ten parts (deciles) based on the model’s predictions, from the highest probability of an event occurring to the lowest. The steeper the plot against deciles, the better the segmentation or lift. We see three lines. The “table” line indicates that the expected mortality is relatively constant across these model deciles, even though the “actual” mortality and the mortality predicted by the “model” vary substantially, indicating significant risk stratification.

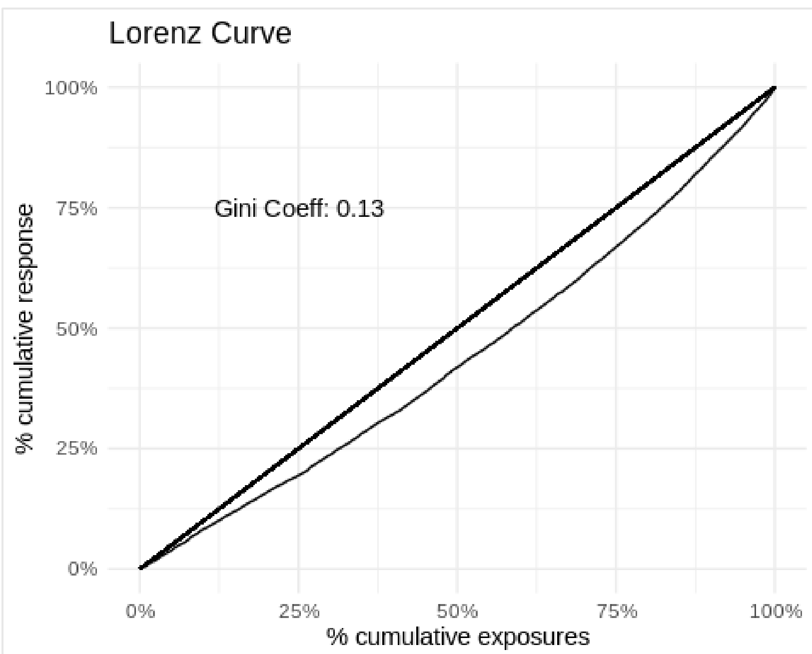
The Lorenz curve described is another way of visualizing the risk stratification of the model. The more bowed the line is from the y=x axis, the greater the Gini coefficient and the greater the risk stratification.

Understanding the behavior of the interactions, as well as gain and cover, can give us some macro insight into what the model is doing. The feature interaction table ranks and demonstrates the most important interactions in the model. ‘Gain’ refers to the improvement in accuracy brought by a feature to the branches it is on, thus indicating the feature is important. ‘Cover’ measures the number of times a feature is used to split the data across all trees regardless of the gain in accuracy achieved. A high gain with a high cover suggests a feature that is very useful across many parts of the dataset.

Lift Curve

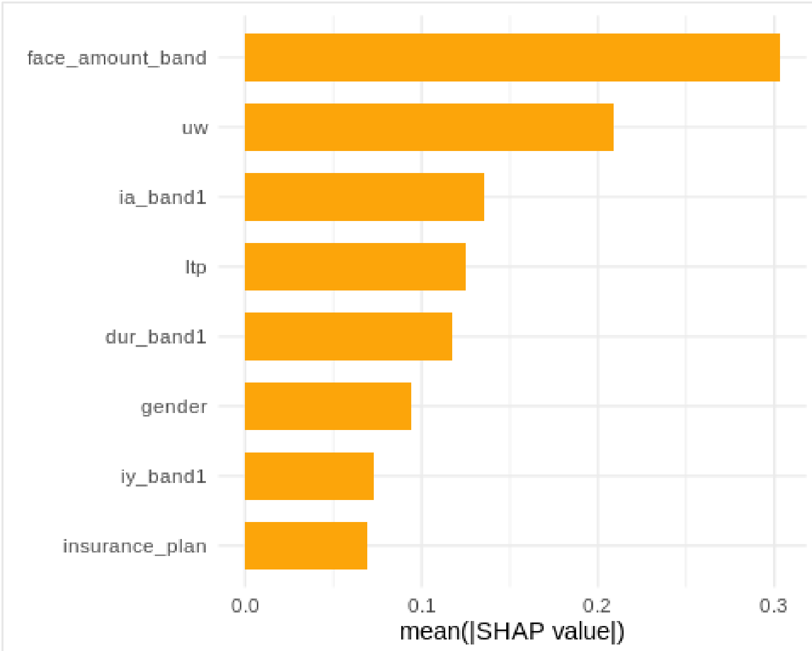


Lorenz Curve



Feature Importance

The following plot is the feature importance plot, which ranks the mean absolute SHAP value for a given feature. It should be noted that being low on the list does not automatically imply that a feature is unimportant. Due to phenomena such as aggregation bias, features with relatively higher numbers of levels can seemingly rank higher than those with lower numbers of levels. Here, the top three tend to have large numbers of levels versus the bottom four.



Feature Interaction Table

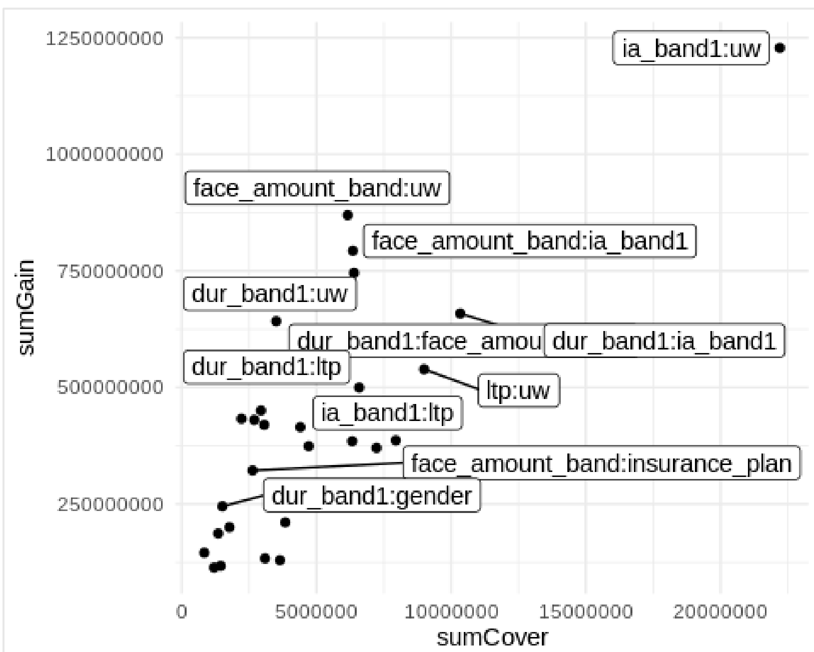
We also develop a table of interaction strengths, sorted by the total contribution to explaining variation in the data. Again, aggregation bias can distort the ranking, so interpreting the ranking should be taken with caution.

Feature	sumGain	sumCover	frequency	meanCover	meanGain
ia_band1:uw	1,228,300,000	22,200,000	3,003	7,393	409,024
face_amount_band:uw	869,600,000	6,159,000	2,112	2,916	411,742
face_amount_band:ia_band1	793,300,000	6,354,000	2,185	2,908	363,066
dur_band1:uw	745,800,000	6,395,000	1,677	3,813	444,723
dur_band1:ia_band1	658,700,000	10,341,000	1,739	5,947	378,781
dur_band1:face_amount_band	642,300,000	3,509,000	1,522	2,306	422,011
dur_band1:ltp	602,600,000	4,825,000	1,101	4,382	547,321
ltp:uw	539,000,000	8,995,000	1,432	6,281	376,397
ia_band1:ltp	499,900,000	6,590,000	1,294	5,093	386,321
insurance_plan:uw	450,500,000	2,941,000	683	4,306	659,590
face_amount_band:gender	432,800,000	2,219,600	1,018	2,180	425,147
gender:uw	430,400,000	2,691,000	1,119	2,405	384,629
gender:ia_band1	420,000,000	3,063,000	1,136	2,696	369,718
face_amount_band:ltp	414,800,000	4,405,000	1,179	3,736	351,824
iy_band1:uw	386,600,000	7,947,000	1,133	7,014	341,218
ia_band1:iy_band1	384,800,000	6,330,000	1,041	6,081	369,645
face_amount_band:iy_band1	374,000,000	4,715,000	1,023	4,609	365,591
ia_band1:insurance_plan	370,400,000	7,233,000	964	7,503	384,232
face_amount_band:insurance_plan	321,900,000	2,630,400	734	3,584	438,556
dur_band1:gender	245,090,000	1,512,500	688	2,198	356,235
dur_band1:iy_band1	210,720,000	3,843,000	631	6,090	333,946

Feature	sumGain	sumCover	frequency	meanCover	meanGain
dur_band1:insurance_plan	200,320,000	1,770,400	529	3,347	378,677
insurance_plan:ltp	186,950,000	1,353,500	164	8,253	1,139,939
gender:insurance_plan	145,750,000	837,300	310	2,701	470,161
insurance_plan:iy_band1	133,400,000	3,095,000	427	7,248	312,412
iy_band1:ltp	129,660,000	3,644,000	556	6,554	233,201
gender:iy_band1	117,310,000	1,447,700	421	3,439	278,646
gender:ltp	113,770,000	1,203,000	471	2,554	241,550

Gain versus Cover

As noted above, 'gain' refers to the improvement in accuracy brought by a feature to the branches it is on, thus indicating the feature is important. 'Cover' measures the number of times a feature is used to split the data across all trees regardless of the gain in accuracy achieved. A high gain with a high cover suggests a feature that is very useful across many parts of the dataset.



4.2.4 FEATURE PLOTS

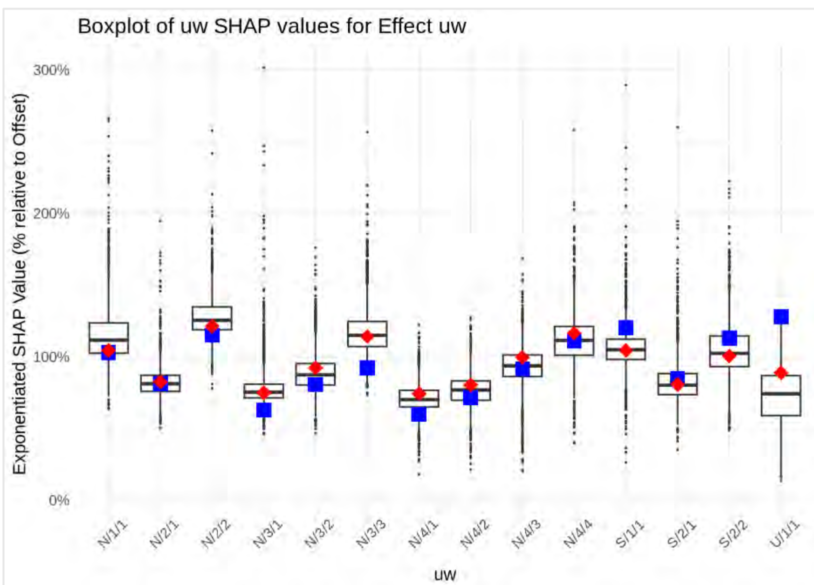
It is useful to plot SHAP values for their main effects (e.g., SHAP values for face amount band by face amount band), as well as interactions (e.g., same, but stratified in some way by other variables). Traditionally, scatter plots are used. However, due to overplotting, it is not clear what is going on with the SHAP values. Here we use boxplots of the SHAP values instead of scatter plotting. This provides a sense of the spread of the SHAP values along with the median and outliers. This is particularly useful for qualitatively evaluating whether there are any meaningful interactions.

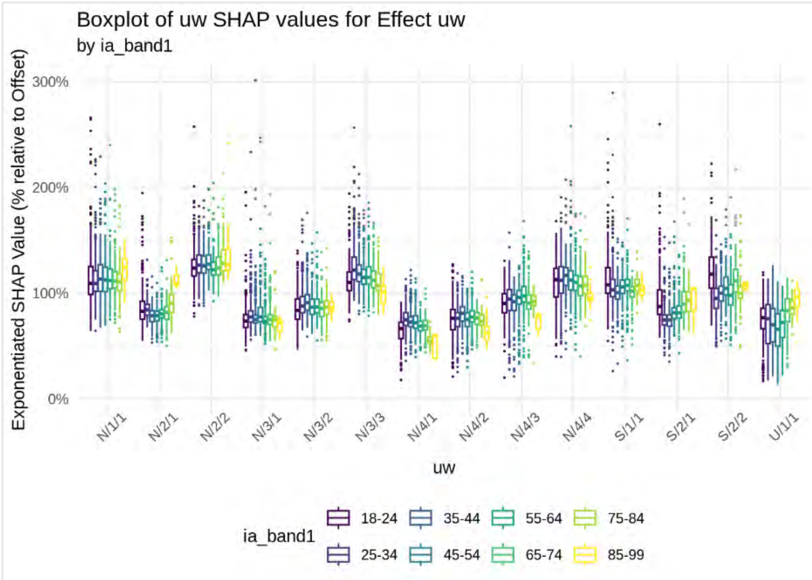
In what follows, red diamonds are mean SHAP values, while blue squares are mean mortality from a subset of the data. Note that SHAP values are partial effects which work in concern with the other features. Therefore, the mean actual mortality will not necessarily be captured by the variability of the feature SHAP values.

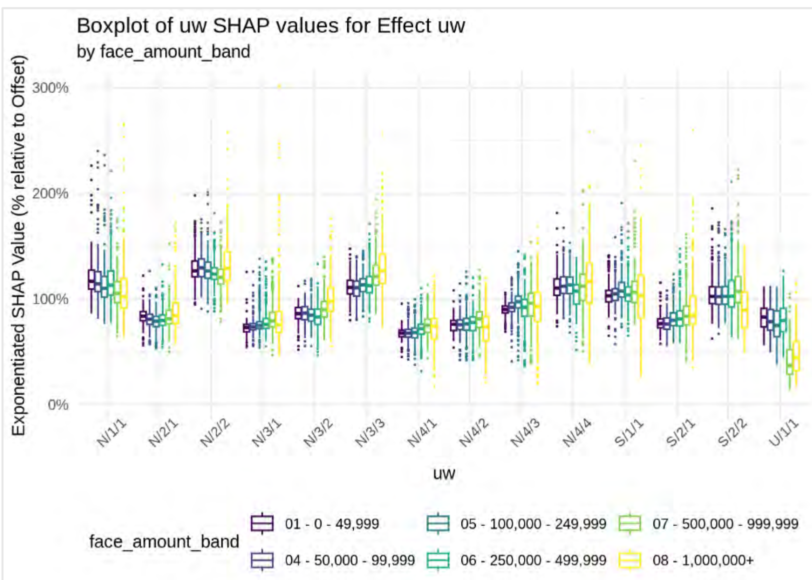
Some patterns are noticeable. We discuss them for each group. You can visually detect an interaction by checking whether the box plots are all on the same level for a given subgroup.

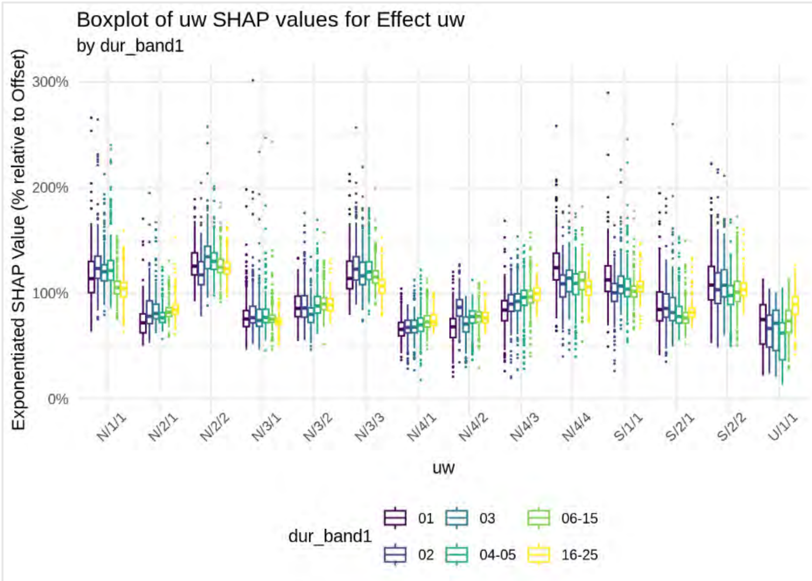
uw: main effect and interactions

- Main effect
  - The spreads from highest to lowest risk classes are similar across non-smoker class systems.
  - Smoker differentiation is narrower than for two-class non-smokers.
- Interaction with face amount band
  - The interaction between underwriting and face amount band, for the underwriting effect, appears confined mostly to three-class non-smoker (N/3/*).
  - Higher face amount bands ($250K+ in the light dataset) appear to have a larger spread of effects.
- Interaction with issue age band: possible narrowing at older ages for two- or four-class non-smokers.
- Interaction with observation year: possible widening of spread of four-class non-smokers with increasing observation year.




Boxplot of uw SHAP values for Effect uw
by ia_band1
Exponentiated SHAP Value (% relative to Offset)
300%
200%
100%
0%
N/1/1
N/2/1
N/2/2
N/3/1
N/3/2
N/3/3
N/4/1
N/4/2
N/4/3
N/4/4
S/1/1
S/2/1
S/2/2
U/1/1
uw
ia_band1
18-24
25-34
35-44
45-54
55-64
65-74
75-84
85-99


Boxplot of uw SHAP values for Effect uw
by face_amount_band
Exponentiated SHAP Value (% relative to Offset)
300%
200%
100%
0%
N/1/1
N/2/1
N/2/2
N/3/1
N/3/2
N/3/3
N/4/1
N/4/2
N/4/3
N/4/4
S/1/1
S/2/1
S/2/2
U/1/1
uw
face_amount_band
01 - 0 - 49,999
04 - 50,000 - 99,999
05 - 100,000 - 249,999
06 - 250,000 - 499,999
07 - 500,000 - 999,999
08 - 1,000,000+

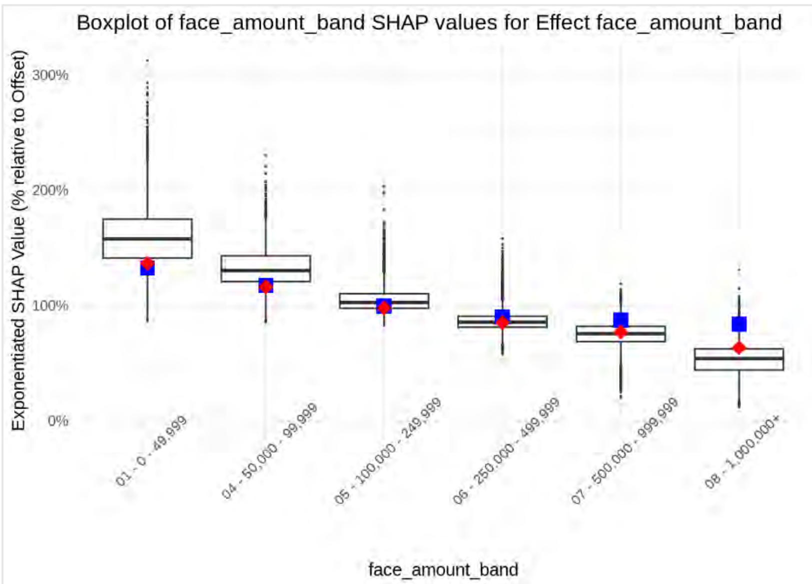


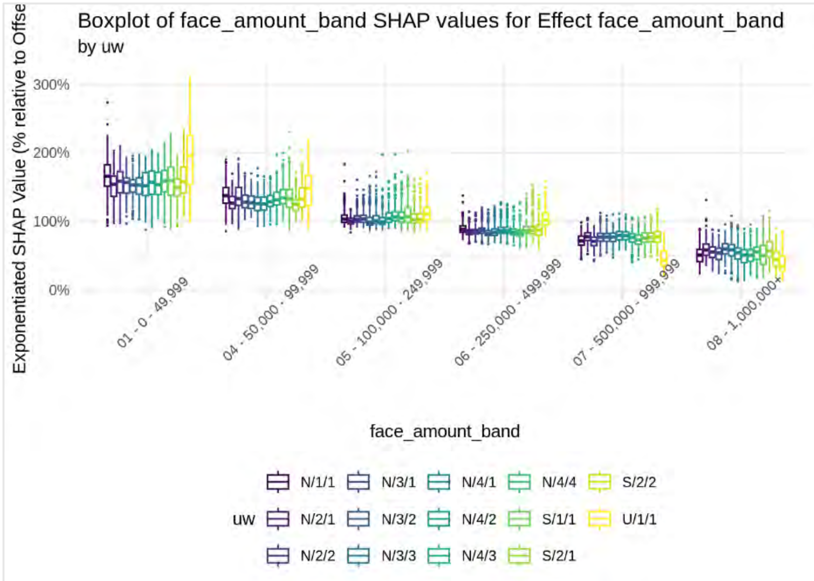
Boxplot of uw SHAP values for Effect uw
by dur_band1
Exponentiated SHAP Value (% relative to Offset)
300%
200%
100%
0%
N/1/1
N/2/1
N/2/2
N/3/1
N/3/2
N/3/3
N/4/1
N/4/2
N/4/3
N/4/4
S/1/1
S/2/1
S/2/2
U/1/1
uw
dur_band1
01
02
03
04-05
06-15
16-25

face_amount_band: main effect and interactions

- Main effect: expected decrease as face amount band increases.
- Interaction with underwriting: face amount effect may be interacting with the Unknown smoker category.
- Interaction with issue age band:
  - o Decreasing effect by issue age for lower bands, flipping to increasing effect by issue age for upper issue age bands
  - o Put another way, spread of face amount effects decreases with increasing issue age

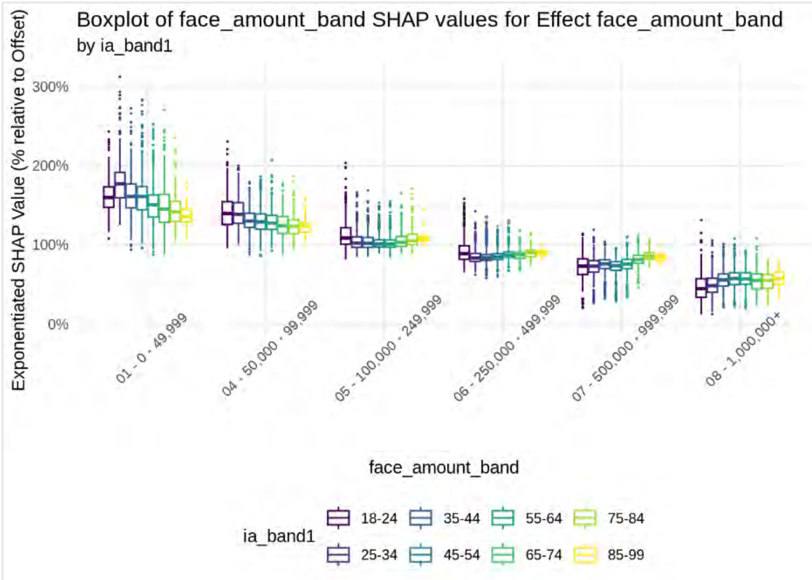
Interaction with Observation Year: no obvious effect



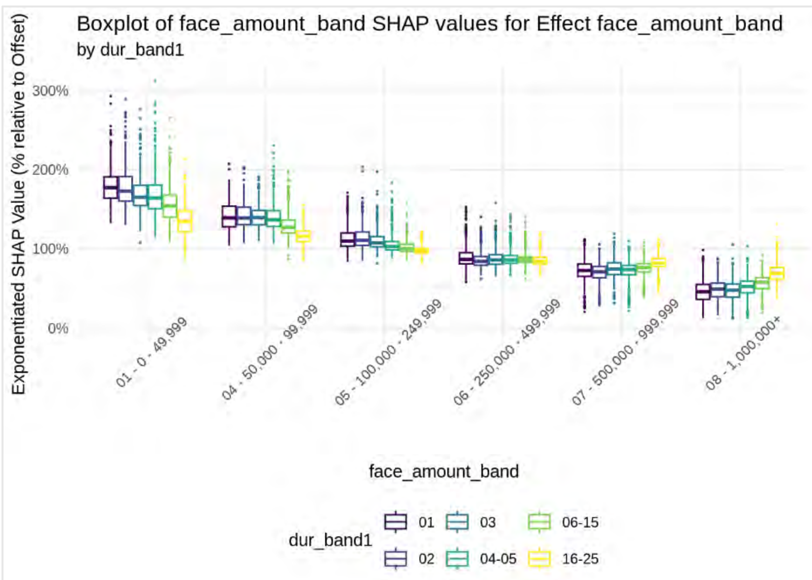


Boxplot of face_amount_band SHAP values for Effect face_amount_band
by uw
Exponentiated SHAP Value (% relative to Offse
300%
200%
100%
0%
01 - 0 - 49,999
04 - 50,000 - 99,999
05 - 100,000 - 249,999
06 - 250,000 - 499,999
07 - 500,000 - 999,999
08 - 1,000,000+
face_amount_band
uw
N/1/1
N/2/1
N/2/2
N/3/1
N/3/2
N/3/3
N/4/1
N/4/2
N/4/3
N/4/4
S/1/1
S/2/1
S/2/2
U/1/1

Boxplot of face_amount_band SHAP values for Effect face_amount_band
by uw
Exponentiated SHAP Value (% relative to Offset)
300%
200%
100%
0%
01 - 0 - 49,999
04 - 50,000 - 99,999
05 - 100,000 - 249,999
06 - 250,000 - 499,999
07 - 500,000 - 999,999
08 - 1,000,000+
face_amount_band
uw
N/1/1
N/2/1
N/2/2
N/3/1
N/3/2
N/3/3
N/4/1
N/4/2
N/4/3
N/4/4
S/1/1
S/2/1
S/2/2
U/1/1

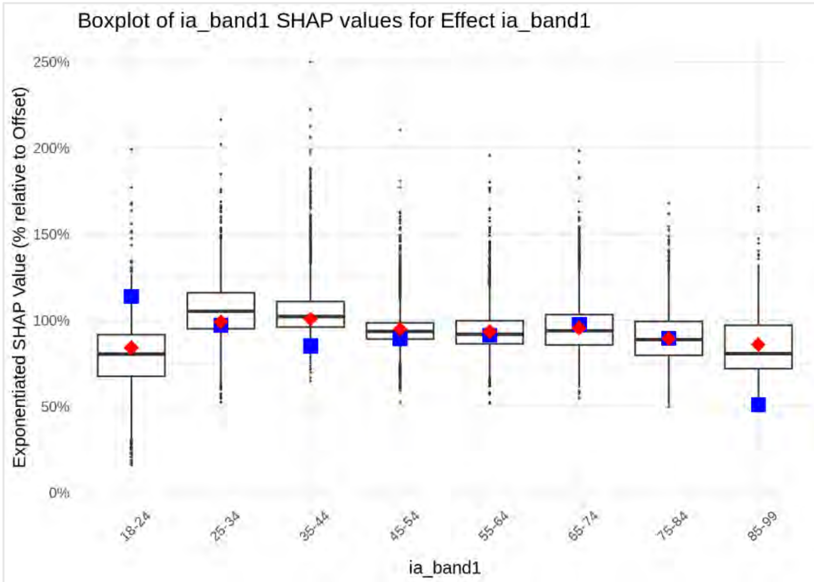


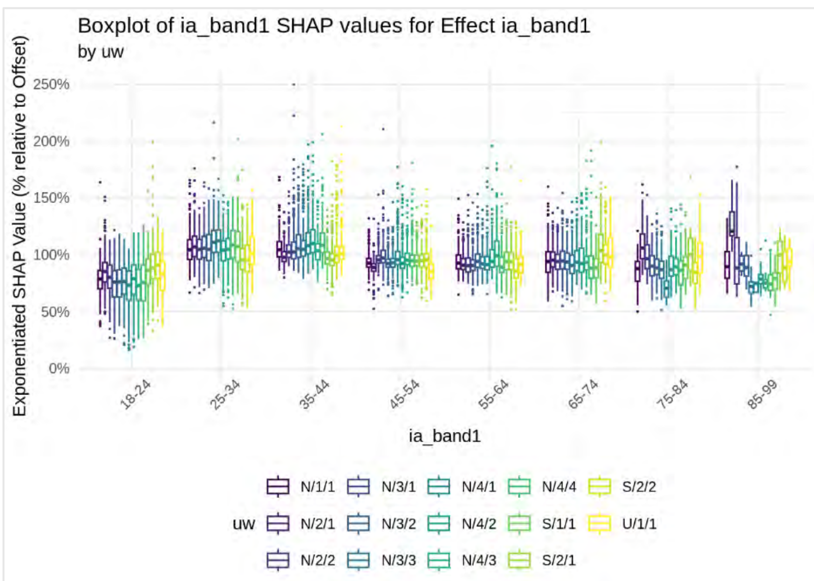
Boxplot of face_amount_band SHAP values for Effect face_amount_band
by ia_band1
Exponentiated SHAP Value (% relative to Offset)
300%
200%
100%
0%
01 - 0 - 49,999
04 - 50,000 - 99,999
05 - 100,000 - 249,999
06 - 250,000 - 499,999
07 - 500,000 - 999,999
08 - 1,000,000+
face_amount_band
ia_band1
18-24
25-34
35-44
45-54
55-64
65-74
75-84
85-99

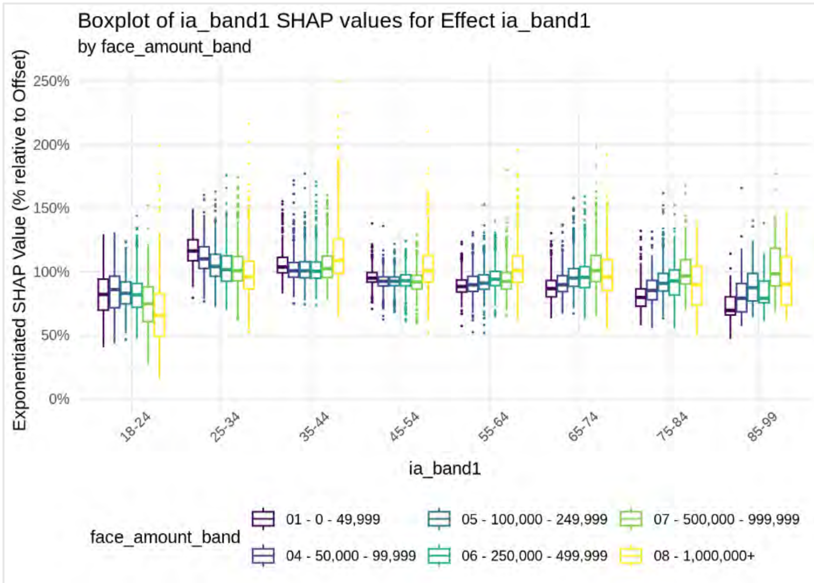


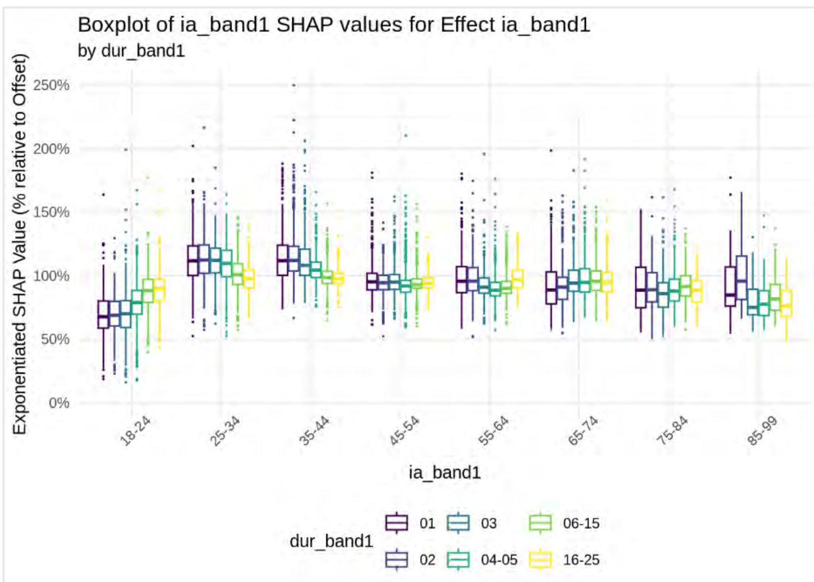
Boxplot of face_amount_band SHAP values for Effect face_amount_band
by dur_band1
Exponentiated SHAP Value (% relative to Offset)
300%
200%
100%
0%
01 - 0 - 49,999
04 - 50,000 - 99,999
05 - 100,000 - 249,999
06 - 250,000 - 499,999
07 - 500,000 - 999,999
08 - 1,000,000+
face_amount_band
dur_band1
01
02
03
04-05
06-15
16-25

ia_band1: main effect and interactions









Issue Age Band (ia_band1)

1. Main Effect: With the exception of ages 18-24, decreasing issue age effect by issue age.
2. Interaction with face amount band: similar to face amount band, spread decreases with increasing issue age.
3. Interaction with underwriting: substantial changes above issue age 75, qualitatively negligible below age 75.
4. Interaction with observation year: no obvious interaction.

4.2.5 GOODNESS-OF-FIT

Goodness-of-fit tables are provided. Each table provides actual-to-model ratios for single variables and for two-way combinations of variables. A model is qualitatively deemed to perform well if goodness-of-fit ratios are close to 100% in almost all situations. The quantitative assessment using significance testing is omitted here.

See the Excel tables in Exhibits 4 and 5 accompanying this report for more information.

4.3 ELASTIC NET GLM

4.3.1 BACKGROUND

Elastic net regularization allows the modeler to combine both LASSO and Ridge penalties into a single model.

As one may recall, ordinary least squares regression requires minimizing the squared difference of the response variable and the predicted values. In symbols,

$$\underset{\beta}{\operatorname{argmin}} \sum_{n} (y - X\beta)^2$$

This is equivalent to maximum likelihood estimation, where one assumes that the response variable y is normally distributed with mean $X\beta$ and variance $\sigma^2 I_{kxk}$. The maximum is taken with respect to β, and the variance parameter is assumed to be fixed but unknown.

The LASSO and Ridge regression methods each add an additional penalty term on the coefficients β. The LASSO adds the sum of the absolute values of the parameters β subject to a tunable weight, λ. This term incentivizes the fitting algorithm to fit toward parameter values close to 0.

$$\underset{\beta}{\operatorname{argmin}} \sum_{n} (y - X\beta)^2 + \lambda \sum_{k} |\beta_k|$$

The Ridge penalty adds the sum of the squares of the parameters β, subject to a tunable weight, α. This term also incentivizes the fitting algorithm to fit toward parameter values close to 0.

$$\underset{\beta}{\operatorname{argmin}} \sum_{n} (y - X\beta)^2 + \alpha \sum_{k} \beta_k^2$$

What may be new to some readers is that, in both cases, for special λ or α, the minimizers of these expressions correspond to the Bayesian maximum a posteriori (MAP) estimators for specific prior distributions for β. In the Ridge case, the prior is the normal distribution with mean 0 and covariance $\tau^2 I_{k\times k}$ for some assumed τ^2.

For the LASSO, the prior is the double-exponential or Laplace distribution with mean 0 and parameter τ.

In either case, it can be shown that if σ^2 and τ are known, the penalizing weights have unique solutions and are equivalent to the k term in Bühlmann credibility. In practice, the penalizing weights are unknown and must be tuned. The resulting optimal β is also credible from a Bayesian perspective. Moreover, it can be shown that these facts carry over to the GLM case.

4.3.2 DATA PREPARATION

Elastic net GLMs as implemented in the glmnet package require that the inputs be converted to model matrices.

4.3.3 MODEL FITTING

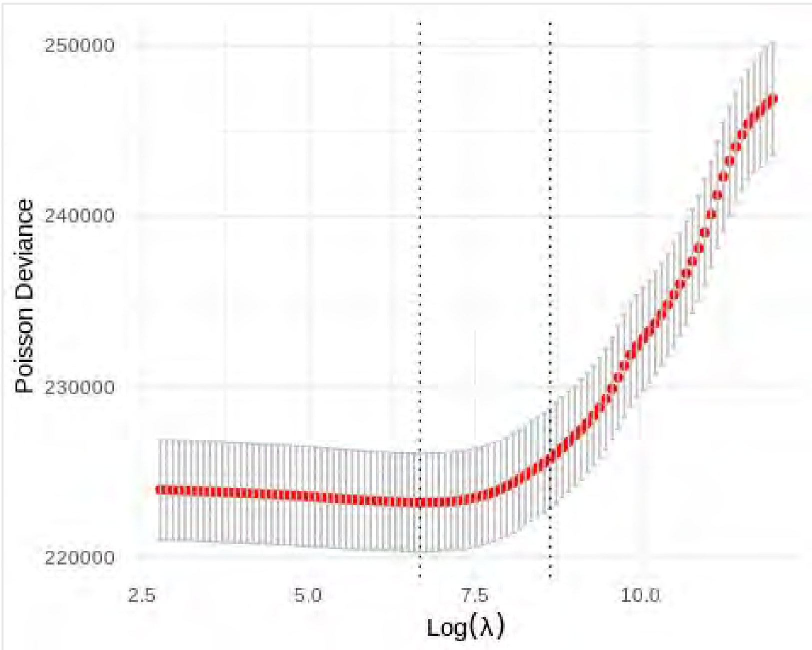
Once the data are set up, we can calibrate a LASSO penalty, lambda, using n-fold cross-validation.

Model Plots and Tables

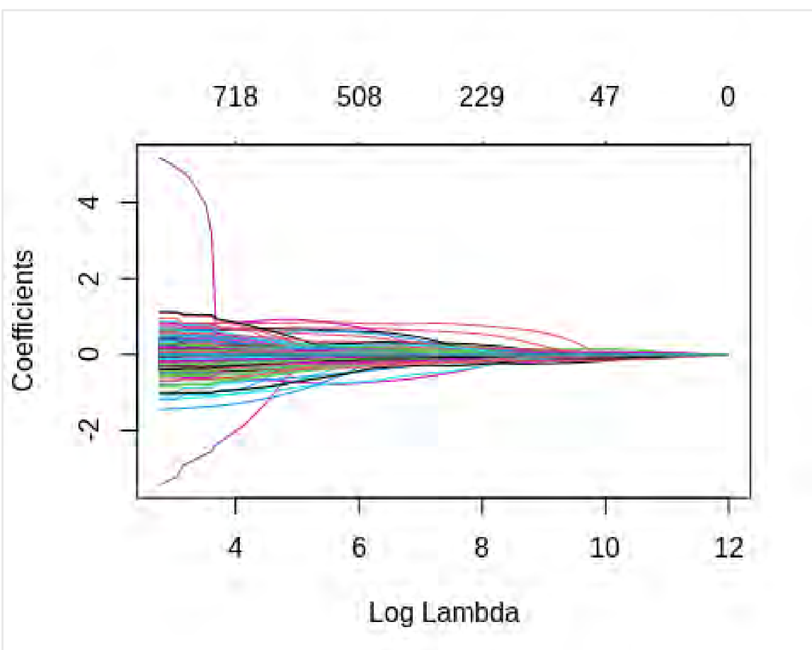
An important by-product of the n-fold cross validation is the plot of λ values. The optimal choice of λ is the lowest, and the model associated with that λ is the final model.

One can also plot the trajectory of coefficients as progressively higher λ impose ever harsher penalties on the coefficients.

Lambda Plot



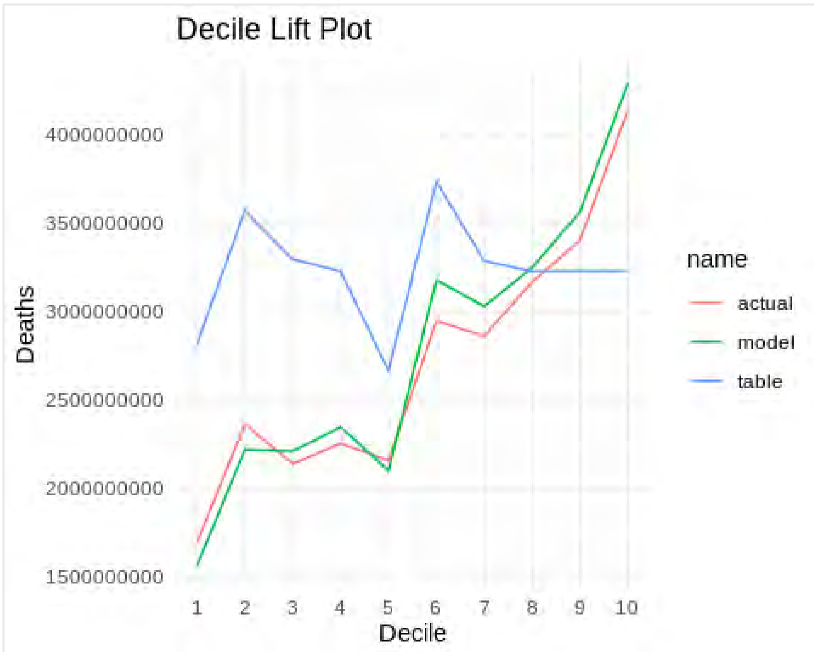
Coefficient Penalization



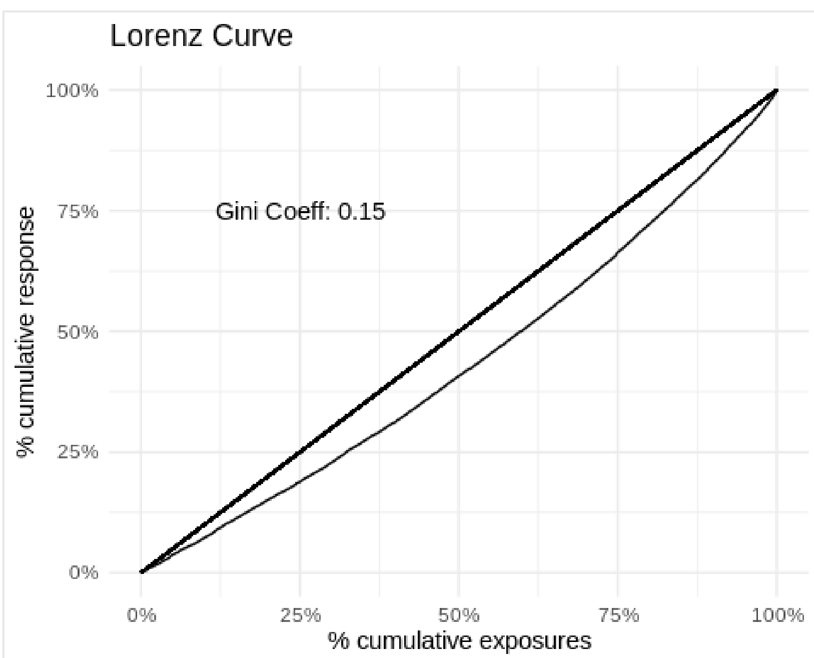
Final Model

The table of coefficients for the elastic net model is substantial. Please see the accompanying Excel file for additional information.

Lift



Lorenz Curve



4.3.4 PLOTS OF TERMS

Plots of the two-way interaction terms, with external factors fixed at their middle values, are provided in Appendix C – Elastic Net Plots of Terms.

4.3.5 TABLES OF TERMS

We provide a table of coefficients from the elastic net model and tables of two-way interactions, with external factors fixed at their middle values.

See the Excel table in Exhibits 6 and 7 accompanying this report for more information.

4.3.6 GOODNESS-OF-FIT

Goodness-of-fit tables are provided. Each table provides actual-to-model ratios for single variables and for two-way combinations of variables. A model is qualitatively deemed to perform well if goodness-of-fit ratios are close to 100% in almost all situations. The quantitative assessment using significance testing is omitted here.

See the Excel tables in Exhibits 8 and 9 accompanying this report for more information.

Section 5: Comparison of Model Predictions

5.1 GOODNESS OF FIT

It is important to compare model performance on the test dataset. Models tend to fit well on the training data.

We compute the MSE, MAD, and Poisson deviance for each model on the test dataset. Models with lower values are considered qualitatively better.

Across all measures, the elastic net GLM model has the lowest deviation, with the LightGBM qualitatively not far behind. The main-effects GLM does not compete, which reinforces the need for some accommodation of interactions.

model	mse	mae	dev
glm	657,407,396,611,204	10,369,070	1,552,142,707,959,196
elasticnet	285,309,606,550,480	8,470,467	723,646,527,740,728
lgbm	355,563,440,625,343	8,741,101	909,129,632,590,165

5.2 GRAPHICAL MODEL COMPARISON

Unlike the GLM, neither the LightGBM nor the penalized GLM provide any information regarding parameter uncertainty. For elastic net GLMs, there are options to estimate parameter uncertainty:

- Move to a fully Bayesian setting. This gives the modeler significant control, at the cost of complexity (e.g., how to choose reasonable priors) and computation cost. Stan and INLA are available for this purpose.
- Apply the method in Tibshirani et al’s “A significance test for the lasso.” This requires rerunning penalized GLMs and is, thus, potentially costly.
- Apply the method in Lederer’s *Fundamentals of High-Dimensional Statistics*, Sec. 5.2. While technically involved, there does not seem to be a heavy computational lift.

To get around the limitations of assessing uncertainty for now, we plot the models versus the envelope of uncertainty arising from the data itself. This shifts the point of view from assessing parameter uncertainty to assessing goodness-of-fit.

The plots are provided in Appendix C – Graphical Comparison of Models.

Broad observations:

- In 2017, some of the average relationships shifted versus 2013-2016. This can be seen by noting the model dots resting outside the error bars.
- The elastic net model may be missing some higher order interactions.

Section 6: Mortality Differences by Product Use Case

6.1 OBSERVATIONS FROM THE RAW DATA

When assessing differences by product, it is not hard to find challenges when looking at the raw, unadjusted data.

One example is that there is virtually no four-class non-smoker exposure in Perm, while there is significant exposure in Term. This implies that there is a potential issue of identifiability in interactions between insurance plan and underwriting due to the imbalance in exposures. This manifests as an apparent instability in calibrations.

For example, the marginal difference between the four-class and two-class non-smokers in the data is 82.42% (78.2%/94.9%), while the marginal difference between Term and Perm is 86.5% (83.4%/96.4%).

No. of Pref. Classes	A/2015VBT
1	97.2%
2	94.9%
3	81.4%
4	78.2%

Insurance Plan	A/2015VBT
Term	83.4%
xL	89.2%
Perm	96.4%
Other	84.1%

By way of comparison, the GLM calibrates 86.8% for Term versus Perm, and the weighted average factors for four-class systems from the GLM model are 79.3% versus 95.7% for two-class, for a ratio of 82.8%. The main effects GLM is, therefore, asserting that both conditions are associated with lower mortality.

For the elastic net GLM, the situation is complicated. All in, there are 52 factors which mention insurance plan, and assessing when perm and term differ is challenging on a bare reading of the factor table.

For the LightGBM model, there is arguably no interesting mean difference for between Perm and Term.

Insurance Plan	Model / 2015VBT
Other	67.0%
Perm	91.3%
Term	93.3%
xL	102.9%

For class system, the LightGBM model is illustrating a substantial reduction in mean mortality for the four-class systems relative to two-class systems.

No. of Pref. Classes	Model / 2015VBT
1	104.4%
2	100.9%
3	101.0%
4	86.2%

All of this strongly suggests the need for more sophisticated analysis.

6.2 OBSERVATIONS FROM THE GLM

One question of interest to actuaries is why different products have different mortality outcomes. Many things could contribute to the difference, such as underwriting practice, anti-selection risk level, market segment, etc., and generally, it is hard to quantify their impact. With the GLM model and relevant analysis, we have a possible solution.

Let us revisit the table output with insurance plan as the predictor of interest.

Weighted Average GLM Factors for Variable: insurance_plan				
	Other	Perm	Term	xL
amount_2015vbt	84.11%	96.38%	83.40%	89.23%
Factor: insurance_plan	100.00%	87.49%	75.96%	95.90%
Ave Fac: dur_band1	93.06%	90.98%	91.02%	90.24%
Ave Fac: face_amount_band	73.00%	79.13%	73.39%	73.66%
Ave Fac: gender	100.67%	100.70%	100.82%	00.60%
Ave Fac: ia_band1	91.24%	92.20%	93.25%	88.70%
Ave Fac: iy_band1	84.52%	91.05%	86.36%	87.41%
Ave Fac: ltp	67.33%	67.33%	84.55%	67.33%
Ave Fac: uw	97.90%	101.27%	90.51%	100.25%

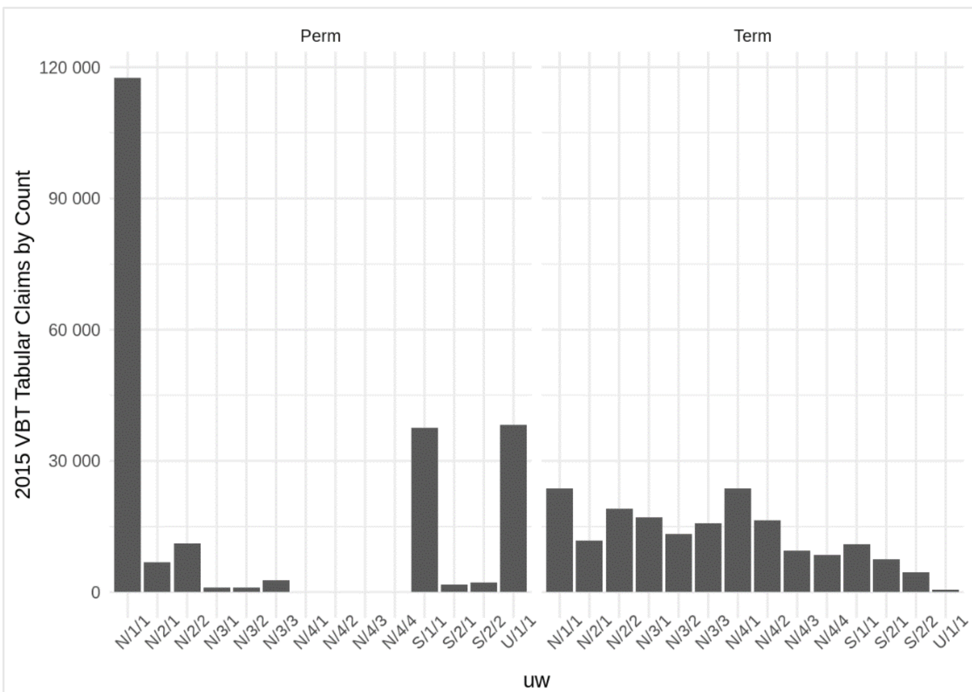
For illustration, let us select Perm and Term for pairwise comparison. By A/15VBT, Perm (96.4%) seems to have worse mortality than Term (83.4%). Is this due to "product differences?"

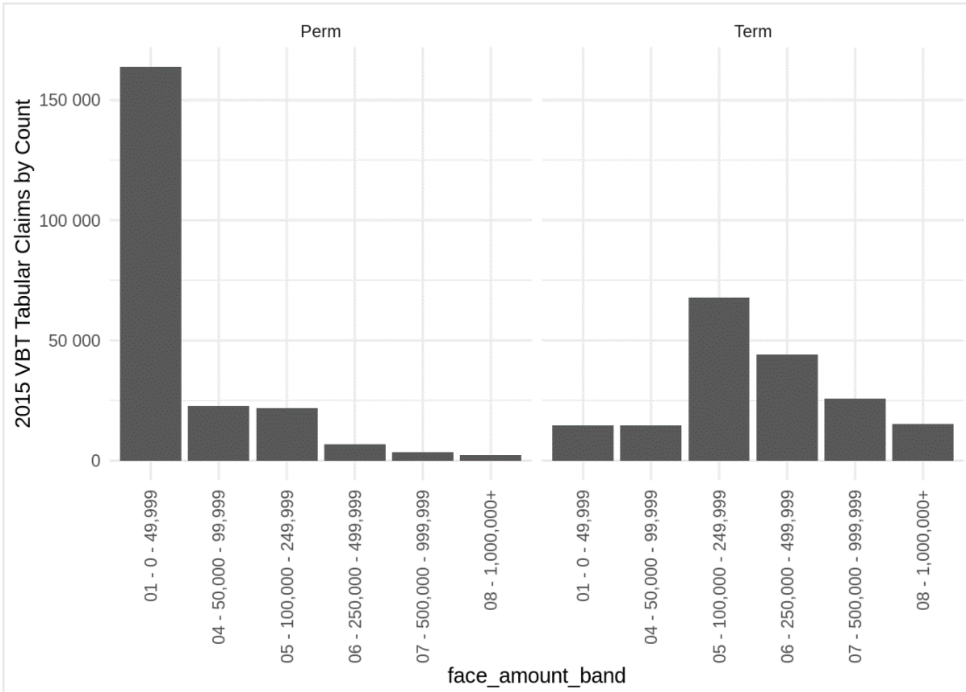
Weighted Average GLM Factors for Variable: insurance_plan in (Perm, Term)		
	Perm	Term
amount_2015vbt	96.38%	83.40%
Factor: insurance_plan	87.49%	75.96%
Ave Fac: dur_band1	90.98%	91.02%
Ave Fac: face_amount_band	79.13%	73.39%
Ave Fac: gender	100.70%	100.82%
Ave Fac: ia_band1	92.20%	93.25%
Ave Fac: iy_band1	91.05%	86.36%
Ave Fac: ltp	67.33%	84.55%
Ave Fac: uw	101.27%	90.51%

Thanks to the GLM model, we can work with a multiplicative formula for prediction. And, with this elegant structure, we can parse out the impact of each individual predictor and make comparisons. The relative impact is represented by the rate of change.

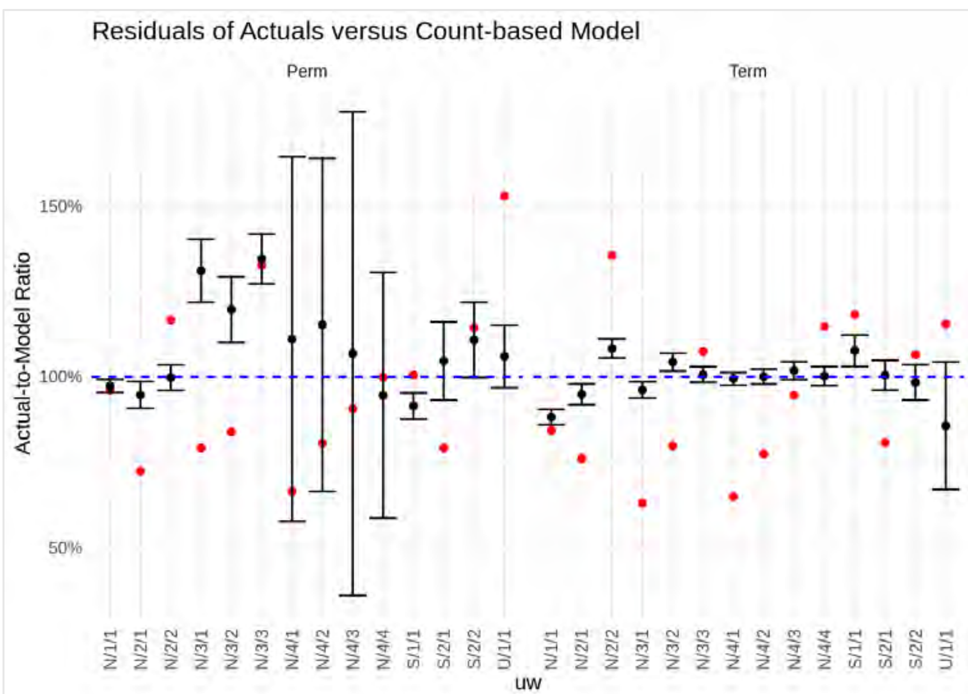
Of these, the movement in underwriting is the most influential, with ratio 111.9% for Perm over Term. This means that, if all the other predictors are controlled, the average underwriting factor on a risk-adjusted basis will make Perm mortality prediction approximately 11.9% higher than that of Term. Other influential drivers from this analysis include face amount band, issue year band, and level term period. This may suggest that, if actuaries/modelers want to build a simpler model yet still capture essential impact to mortality outcome, they might consider including at least those predictors in the GLM model.

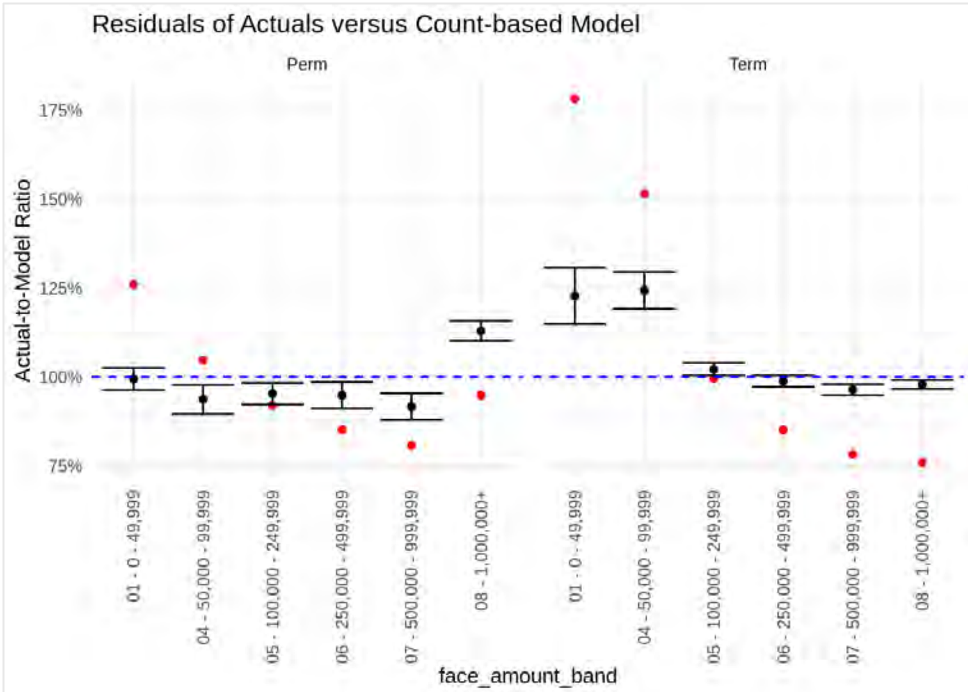
One should nonetheless look to residuals and distributions to ensure that valuable interactions are not being lost. Part of what we are seeing has to do with the different distributions between Perm and Term of underwriting and face_amount_band. Perm tends to favor one- and two-class risk class systems, while Term tends to favor three- and four-class systems. Perm also tends to favor lower face amounts, while Term favors higher face amounts.





In light of what we see for distribution, it is unsurprising that the model fits poorly for the Term subset for smaller face amounts and one- and two-class systems, while the model fits the Perm subset poorly for three- and four-class systems. Since Perm tends to dominate the lower face amounts, model fit is not nearly as poor there as for the Term subset.



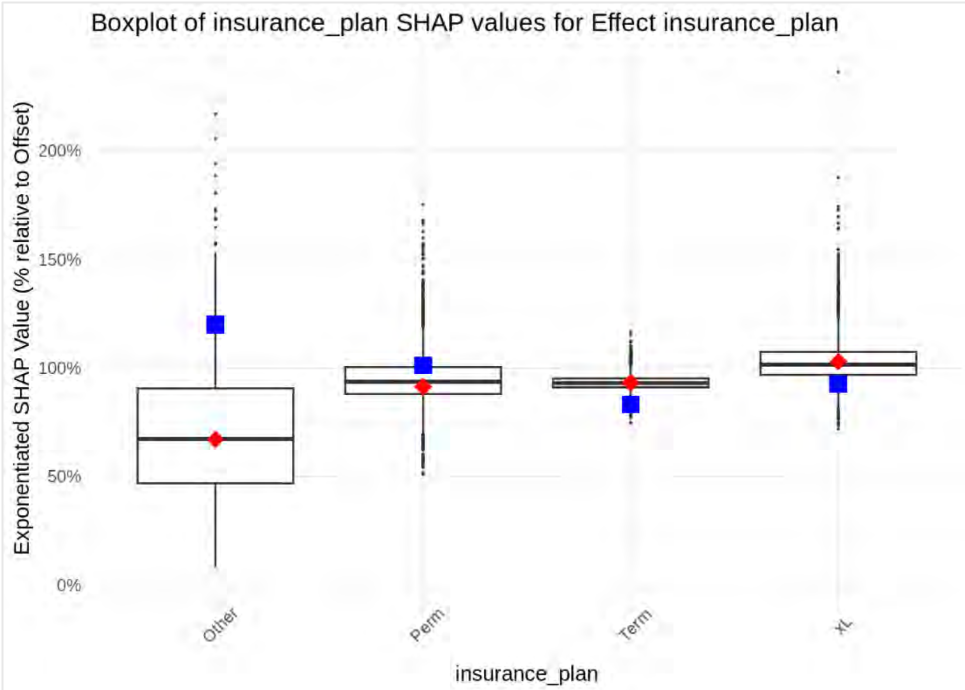


Fitting a main-effects GLM on a dataset is a frequently used first step when modeling any dataset. Analyzing residuals from this model and assessing parameter variability by subgroup can reveal useful patterns for further analysis. It is often the case that interactions of effects are present. While a useful starting point, main-effects models cannot capture such interactions effectively. It is, therefore, necessary to turn to richer models and approaches.

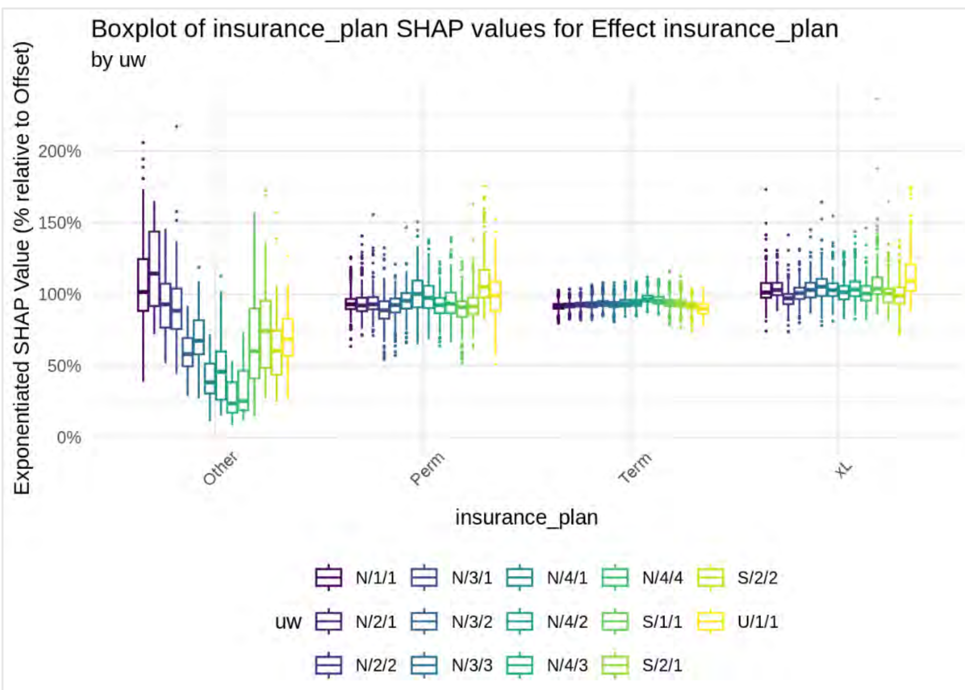
6.3 OBSERVATIONS FROM THE GRADIENT BOOSTING DECISION TREE

Below are box plots of SHAP values for insurance_plan by the other variables.

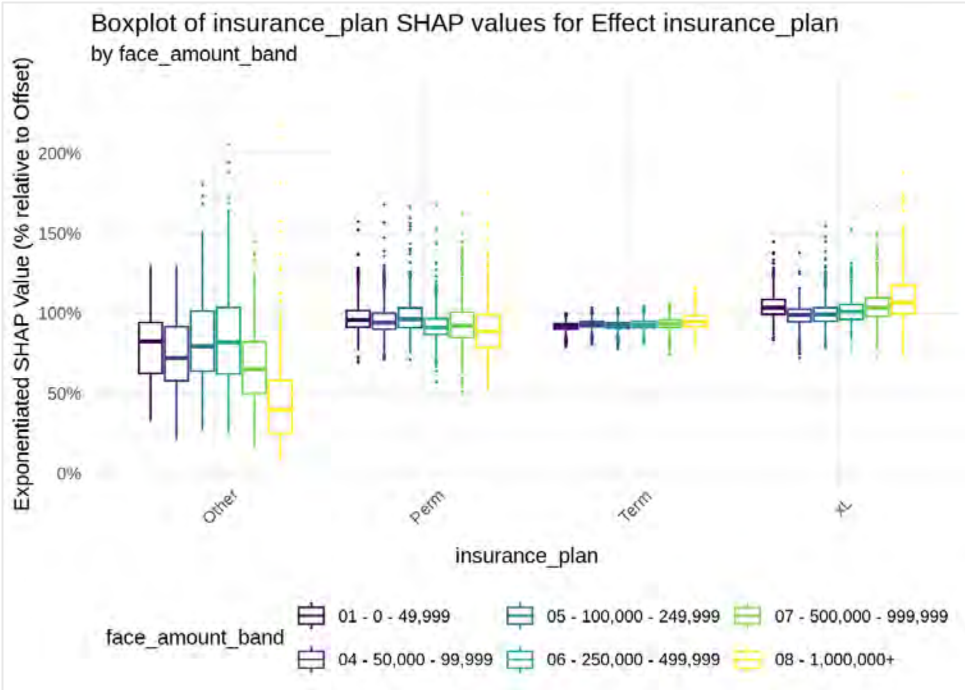
We have noted the interactions with insurance plan from the LightGBM SHAP values as follows:



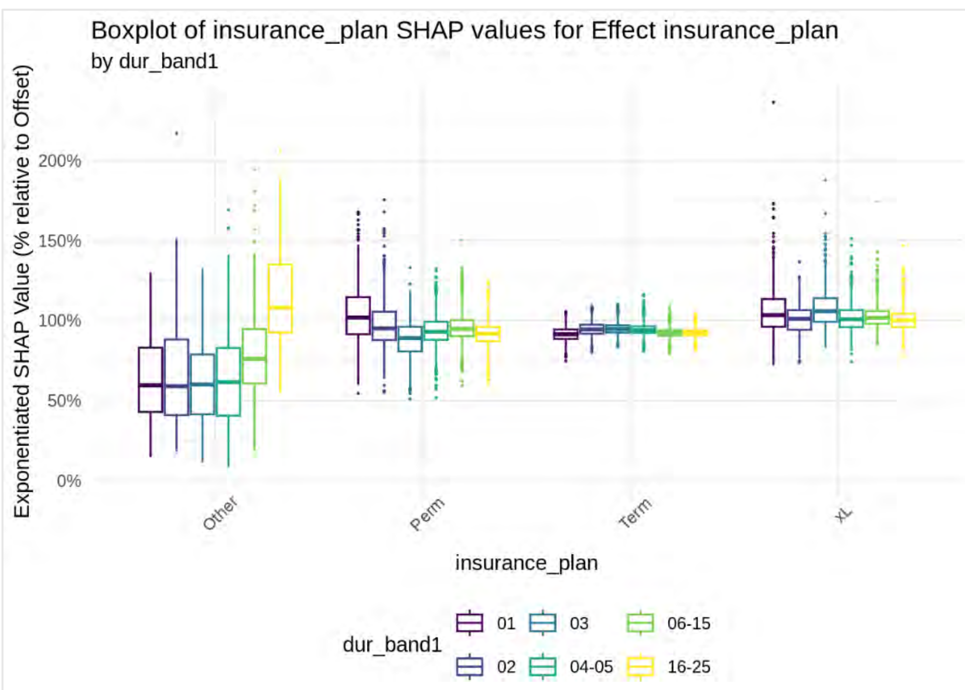
- Main Effect: Perm and Term SHAP distributions are qualitatively similar, with the xL class higher.



- Interaction with underwriting:
  - o Substantial interaction with the "Other" category
  - o For Term, some evidence of higher mortality for N/4/3 and N/4/4

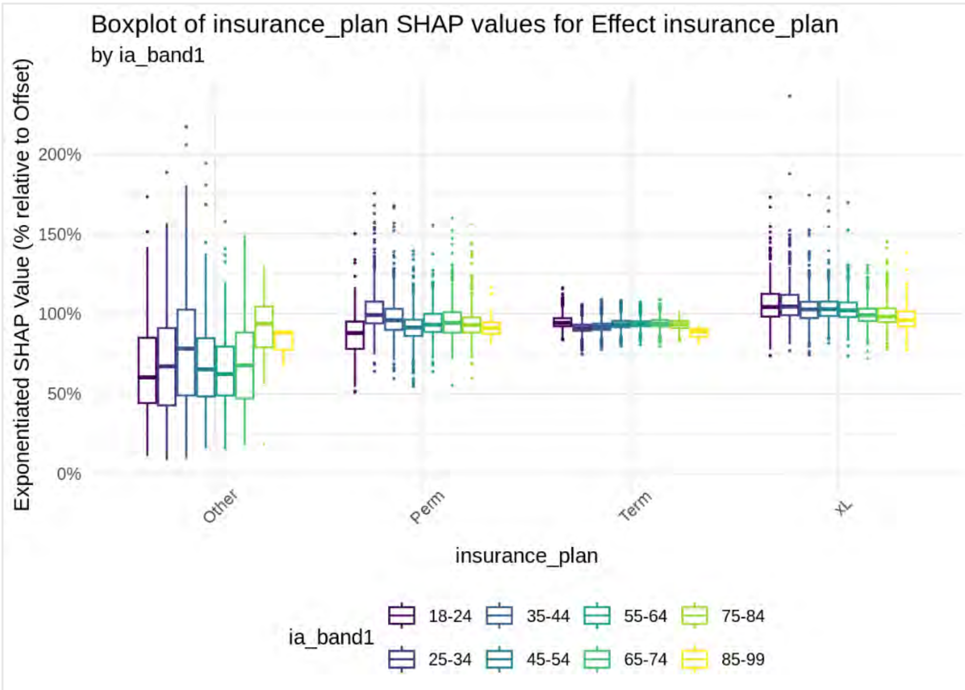


- Interaction with face amount band:
  - No obvious interactions with Perm and Term
  - Weak evidence for interaction with xL, based on U-shaped pattern in boxplots

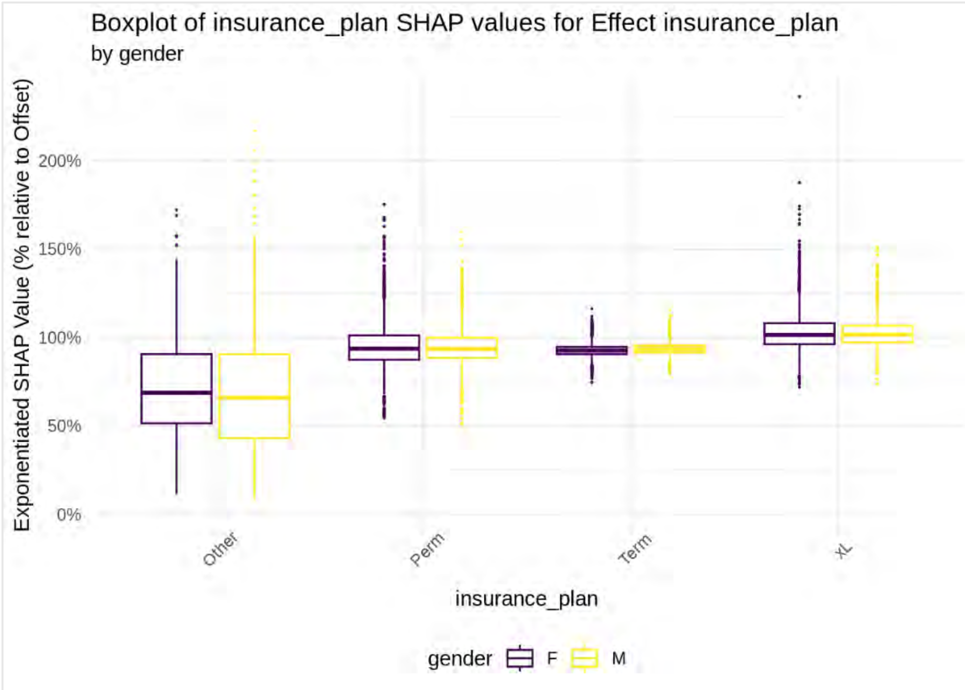


- Interaction with duration
  - Weak evidence for elevated mortality in early durations for Perm
  - Weak evidence for opposite in early durations for Term

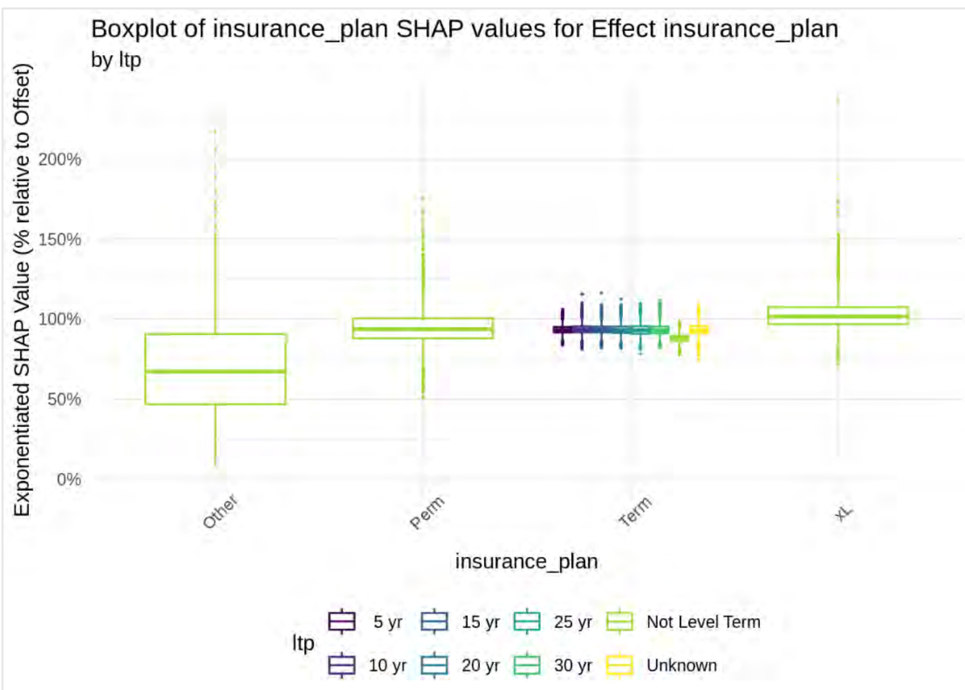
- Face amounts $1 million and higher for “Other” are plainly different from lower face amount “Other”



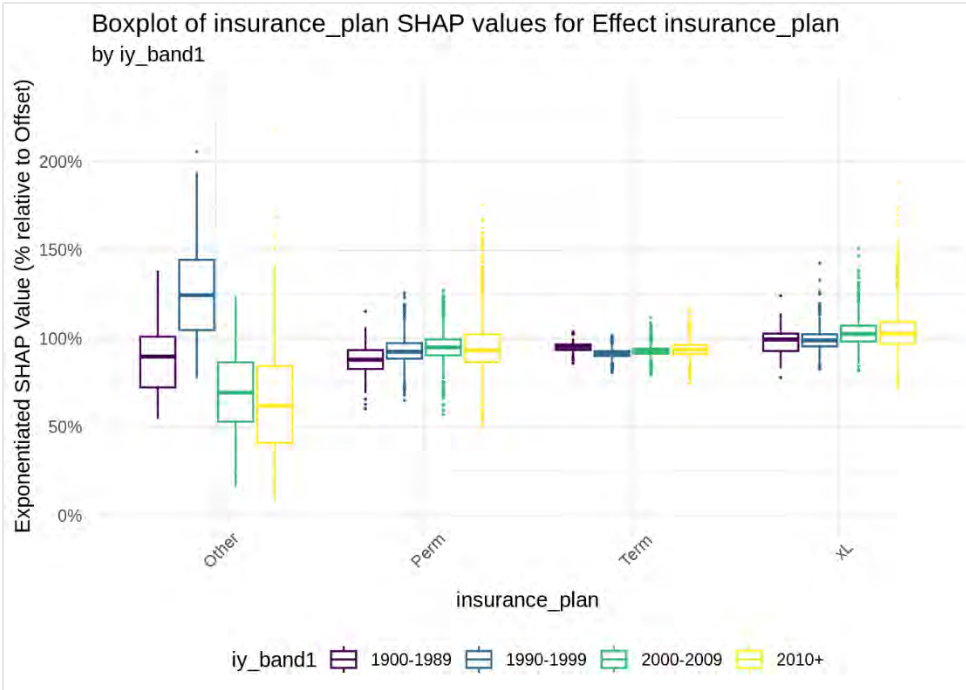
- Interaction with issue age
  - Evidence for different issue age slope (relative to 2015VBT) for xL based on downward trend in boxplots
  - Weak evidence for slight upward issue age slope (relative to 2015VBT) for Term based on trend in boxplots



- Interaction with gender: no obvious interaction



- Interaction with level term period:
  - Obviously, no interactions outside of Term
  - Within Term, “Not Level Term” has lower mortality than the other level Term types

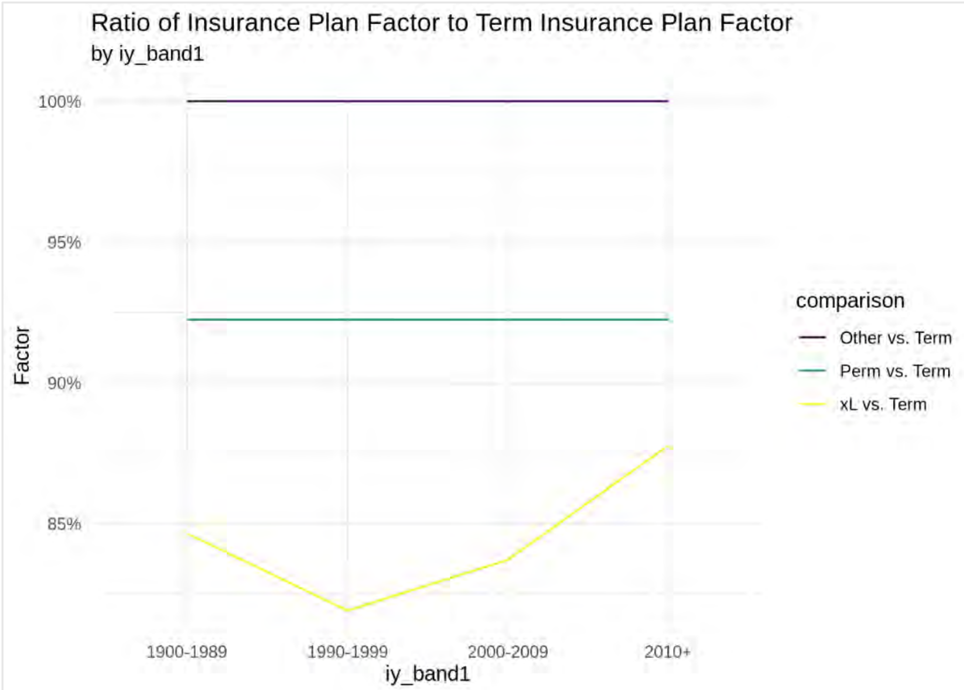


- Interaction with issue year band:
    - Since 1990, there is evidence of an upward trend in mortality for all categories outside of "Other."

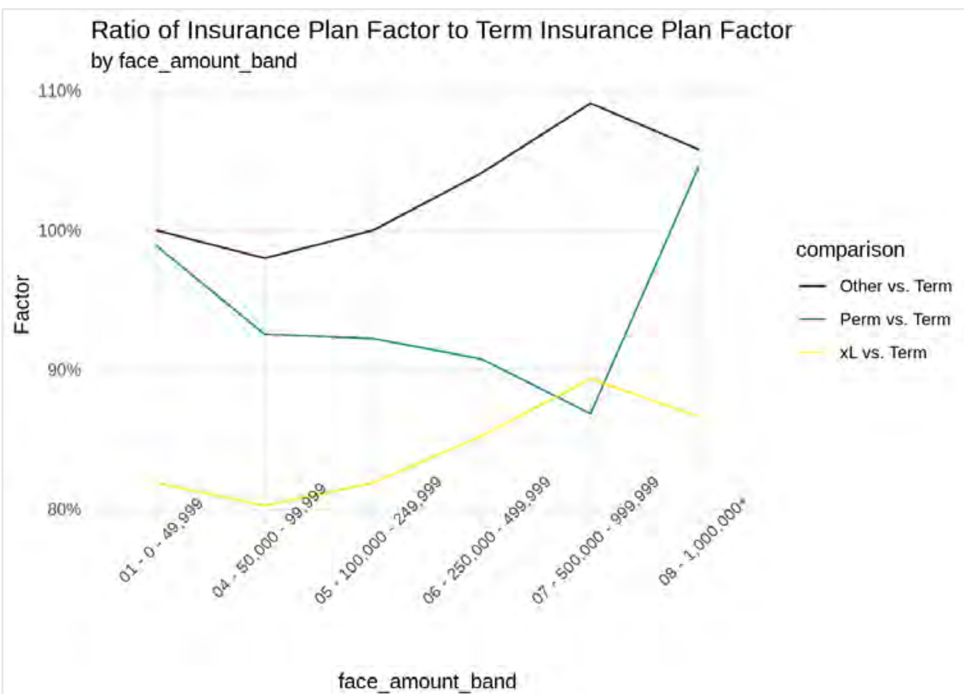
6.4 CONTRASTS OF INTERACTIONS WITH INSURANCE PLAN FROM THE ELASTIC NET MODEL

The elastic net model encodes interesting interactions of insurance plan with other predictor variables. Graphing the contrast between insurance plan types can reveal patterns which are difficult to see when looking at the bare coefficients or tables of factors.

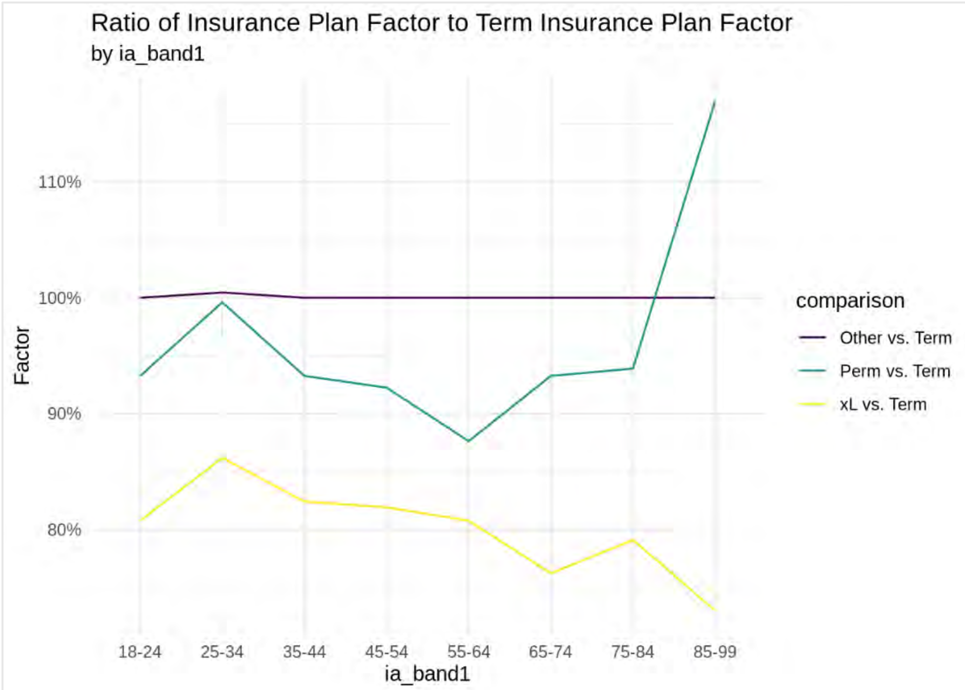
To do so, we can gather the table of factors, which include insurance plan, and compute the ratio of the factors versus the factors for Term. For example, if the marginal factor for Perm males is 99%, and the factor for Term males is 90%, then the ratio is 110%. These contrasts can then be plotted for both males and females across the plan comparisons, as can be seen in the following graphs.



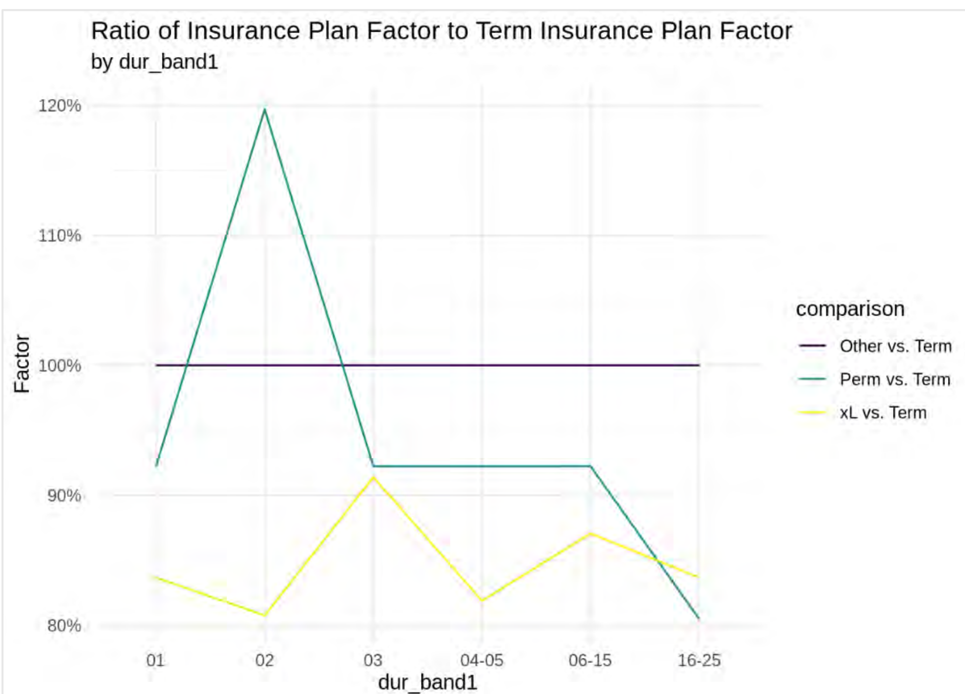
- It appears that the gap between UL/VL/ULSG/VLSG and Term has been narrowing with increasing issue year.



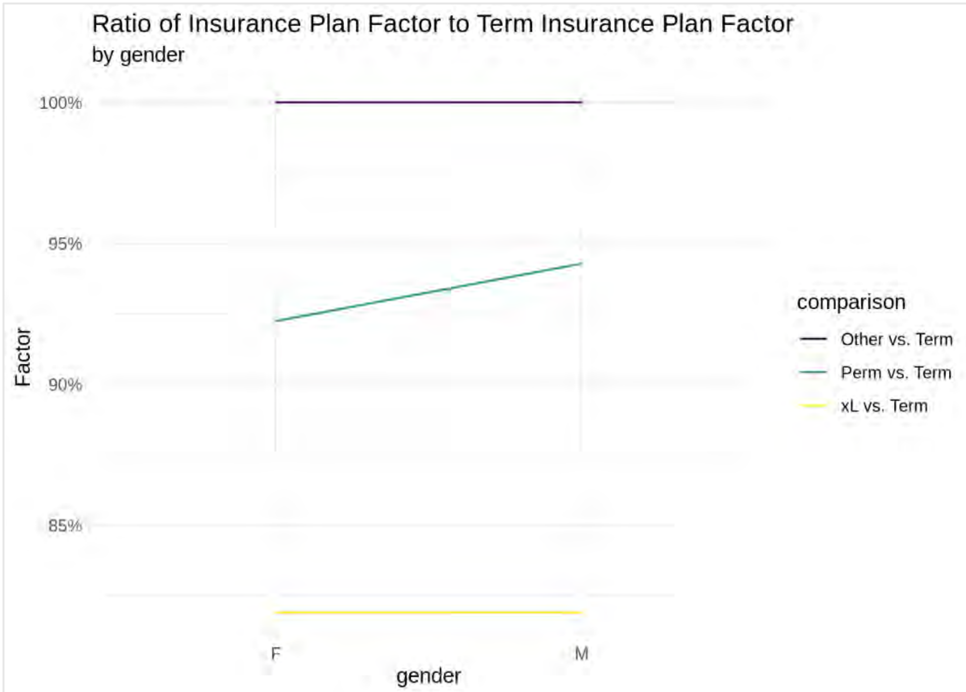
- xL and Other tend to have a wider spread of factors for face amount than Term, while Perm has a narrower spread of face amount factors than Term.



- Perm and xL tend to have a flatter slope than Term by issue age, except above issue age 65. Above issue age 65, the slope of Perm and xL diverges.



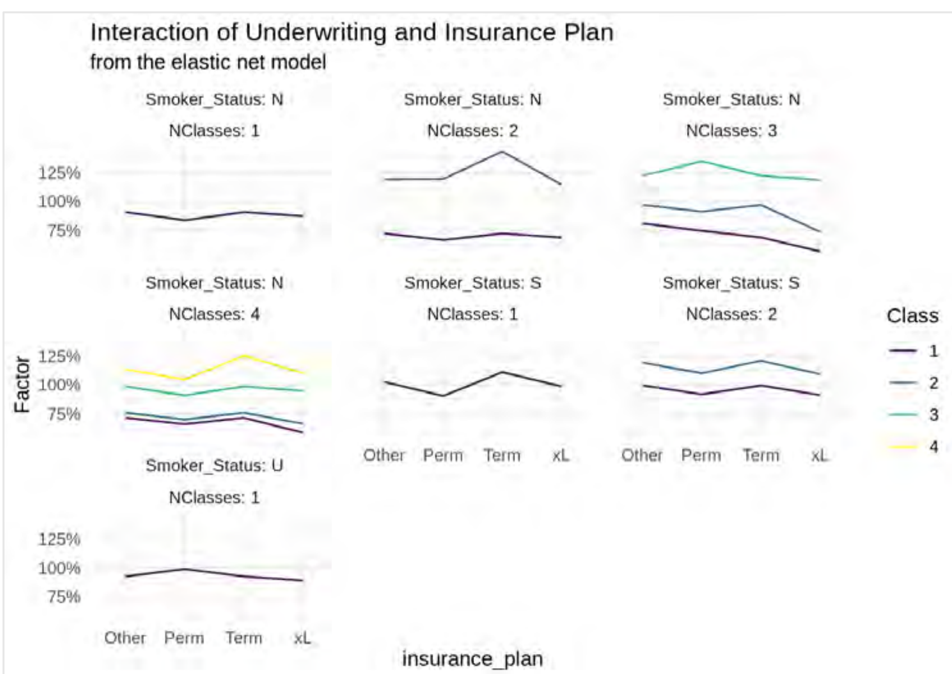
- Perm tends to have higher duration 2 experience than others.



- The gender differential for males is narrower for Perm than for Term.

A different view helps illustrate the interactions of underwriting and insurance plan. It can be seen that

- The residual standard class of a two-class non-smoker system for Term is much higher than the others.
- The spread for Term and xL in the four-class non-smoker system is wider than for Perm and Other.



Section 7: Summary

In this analysis, we explored the application of a predictive modeling framework within actuarial experience studies, focusing on mortality differences by product type in the ILEC dataset. Our analysis revealed several key insights and mortality differentials that can be useful to understand drivers of mortality. This framework and the findings demonstrated the use and considerations required of predictive modeling and underscore its value in making informed actuarial decisions.

Section 8: Acknowledgements

The SOA extends its gratitude to this project subgroup of the Individual Life Experience Committee (ILEC). The project subgroup designed the project, completed/oversaw the analyses, and authored and peer reviewed the report. The members are:

Philip Adams, FSA, MAAA, CERA (chair)

Cynthia Edwalds, FSA, ACAS, MAAA

Brian Holland, FSA, MAAA

Michael Niemerg, FSA, MAAA

Haofeng Yu, FSA, MAAA, CERA

The authors would also like to thank Ed Hui, FSA, MAAA and Steve Ekblad, FSA, MAAA for their contributions throughout the development of the framework.

At the Society of Actuaries Research Institute:

Korrel Crawford, Senior Research Administrator

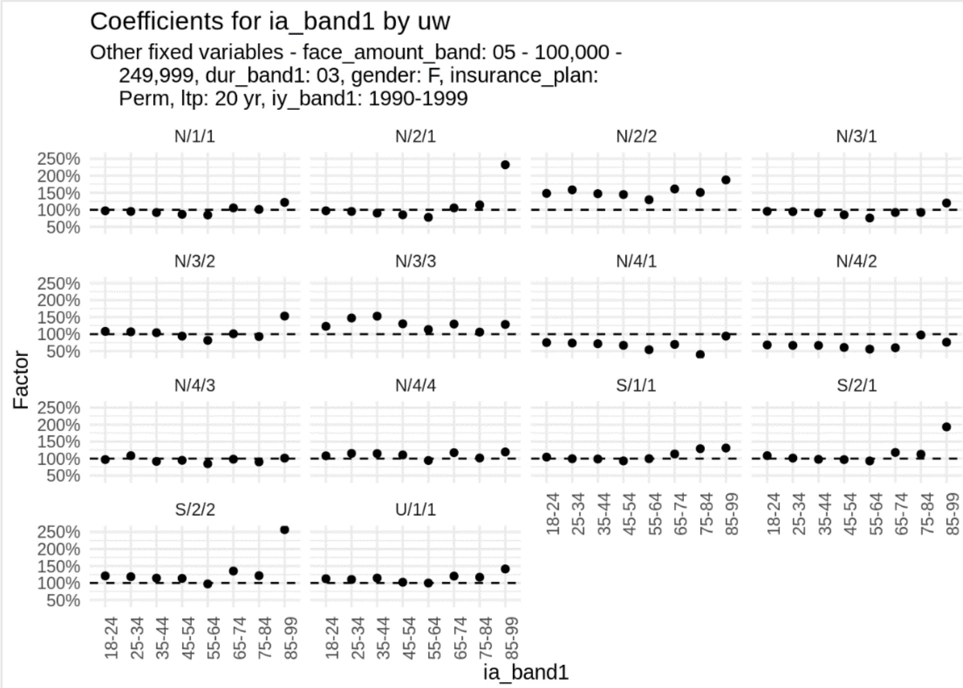
Pete Miller, ASA, MAAA, Experience Studies Actuary

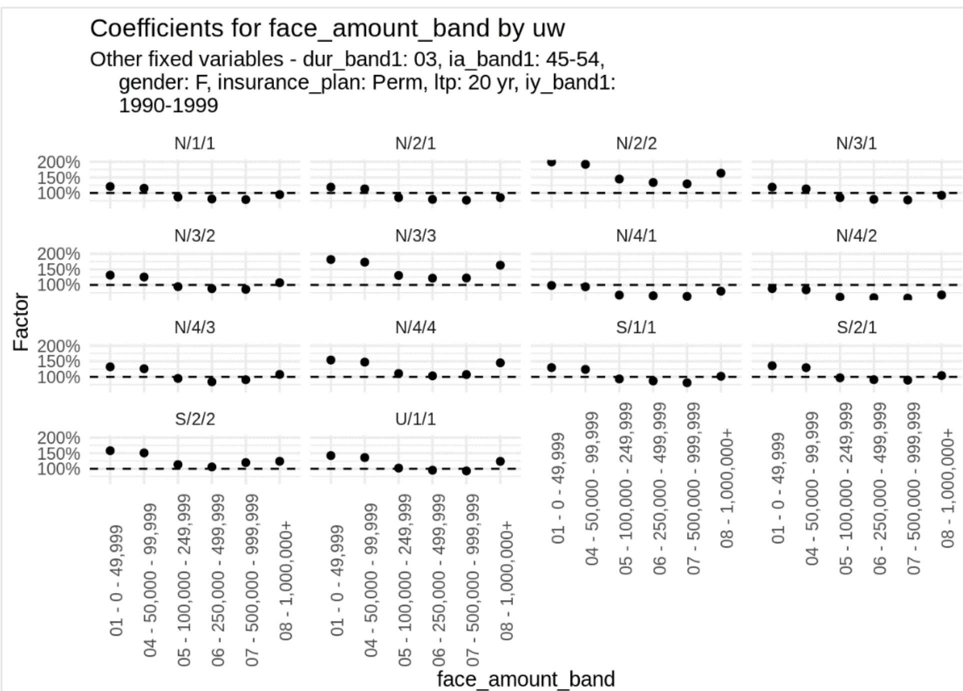
Appendices

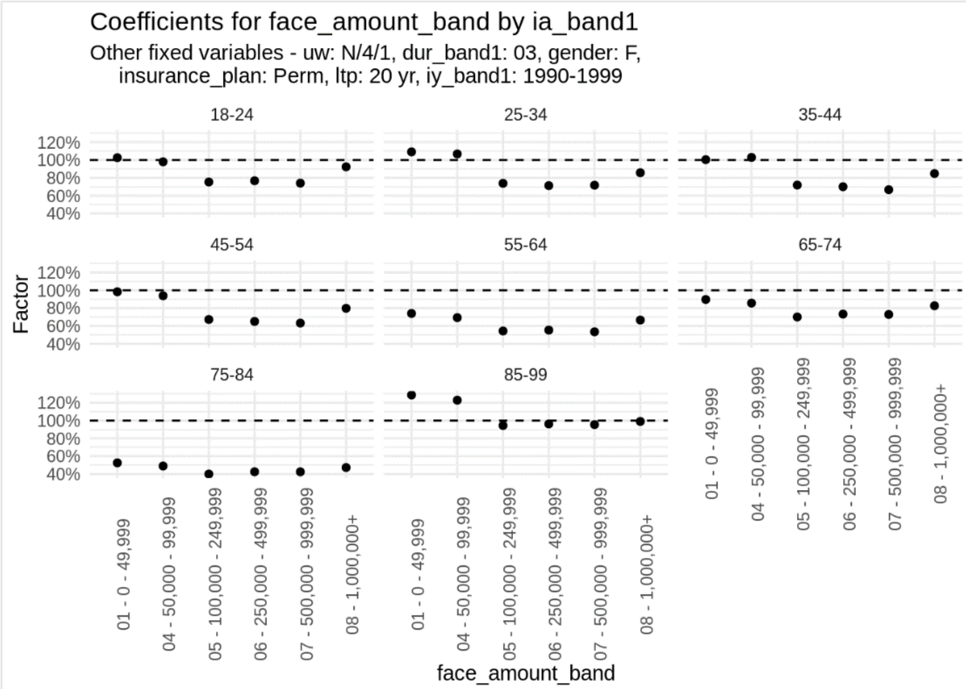
APPENDIX A – FRAMEWORK PARAMETERS

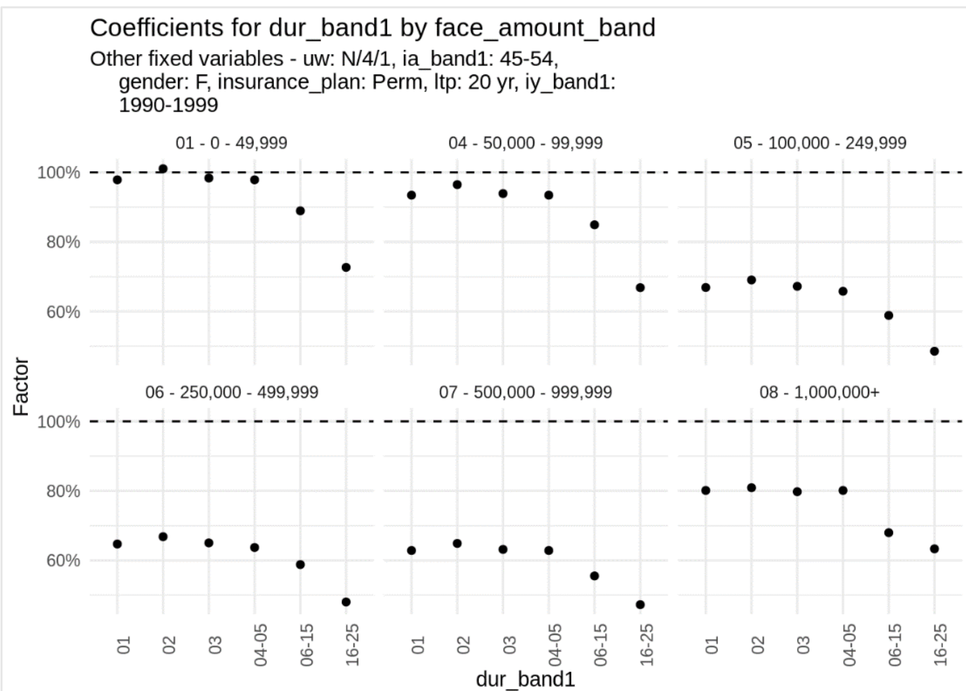
Parameter	Description
prototype	Sets whether to limit the dataset size
prototype_size	Size of the subset from the dataset, if prototype is True
nTrainSeed	Random number seed for generating a train-test split
nGLMNetCores	Number of cores which should be provisioned for elastic net cross validation
nInteractionDepth	How many interactions to include in the elastic net GLM if desired; used if nUseTopLightGBMInteractions is set to 0
fGLMNetAlpha	The weight used in the elastic net GLM for the alpha parameter
nUseTopLightGBMInteractions	Number of two-way interactions from the LightGBM analysis to use in the elastic net GLM
bUseSparse	Instruct the glmnet package to use sparse matrices
nELSeed	Seed for elastic net cross validation
flgbm_vis_subset	Random fraction of data to use for LightGBM ridgeplot visualizations
bFullInteractions	Requests full interaction SHAPs
nPlotTopFeatures	How many top features to plot for LightGBM
nPlotTopInteractions	How many top interactions to plot for LightGBM
nGBMSeed	Seed for LightGBM
runGLM runLightGBM runGLMInt	Switches to include the subanalyses for each model
bDebug	For code that relies on it, set to debugging
bUseCache	Large or difficult-to-compute objects are saved; if True, saved objects are loaded if bInvalidateCaches is also False
bInvalidateCaches	Set to True to force computation of all objects
src_file	Source file for data, can be either CSV or Parquet format, and local or HTTP hosted. If AWS, it will detect this situation and use Arrow's API to load the data
resp_var	Name of the response variable
resp_offset	Name of the response offset
pred_cols	Name of the predictor columns
factor_cols	Name of the columns to treat as factors

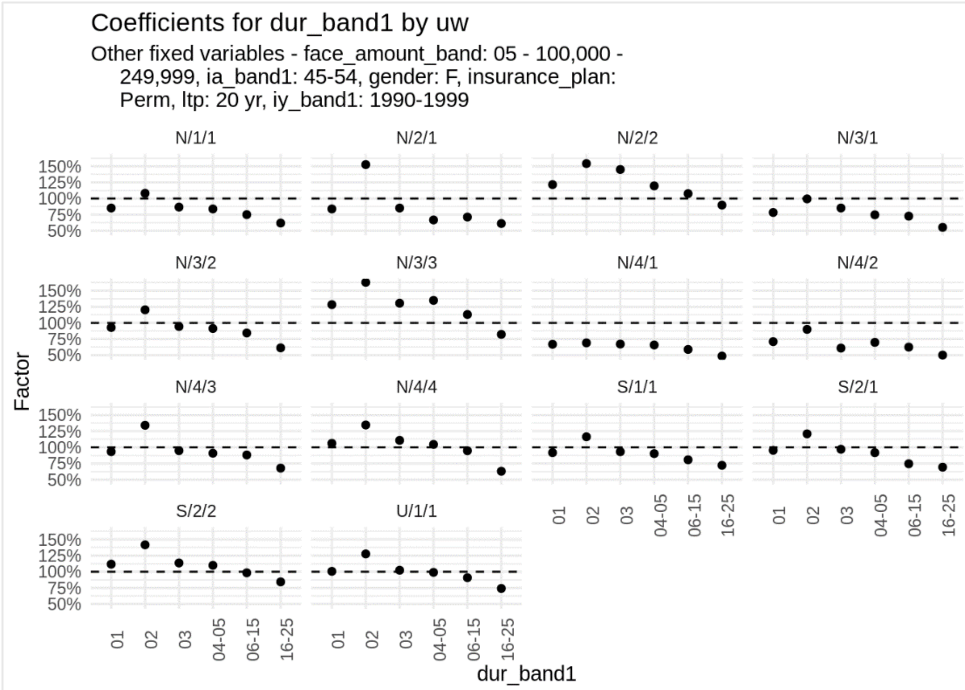
APPENDIX B – ELASTIC NET PLOTS OF TERMS

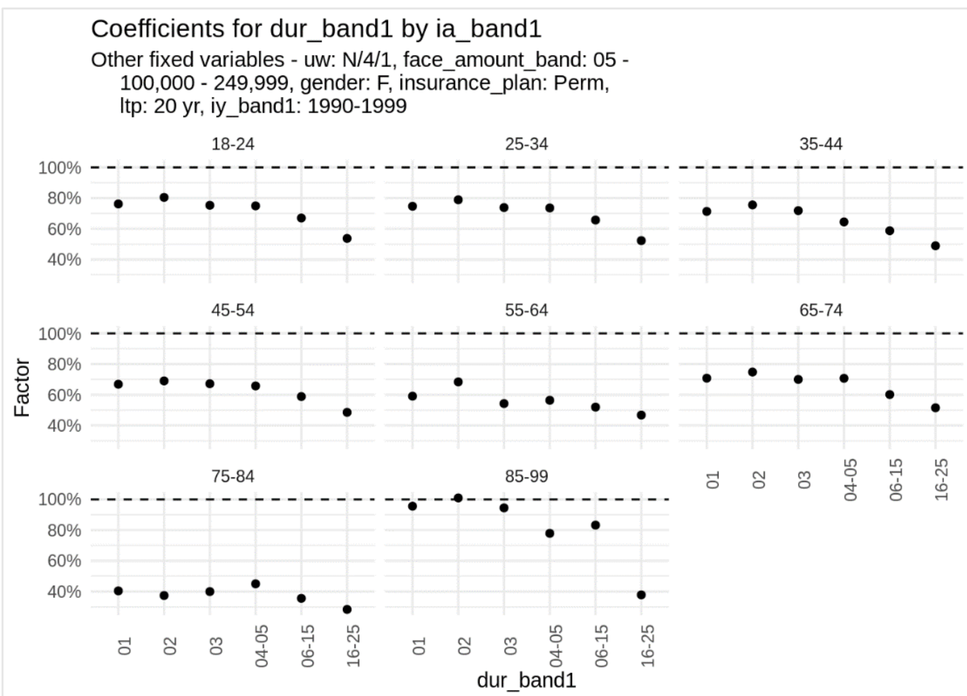


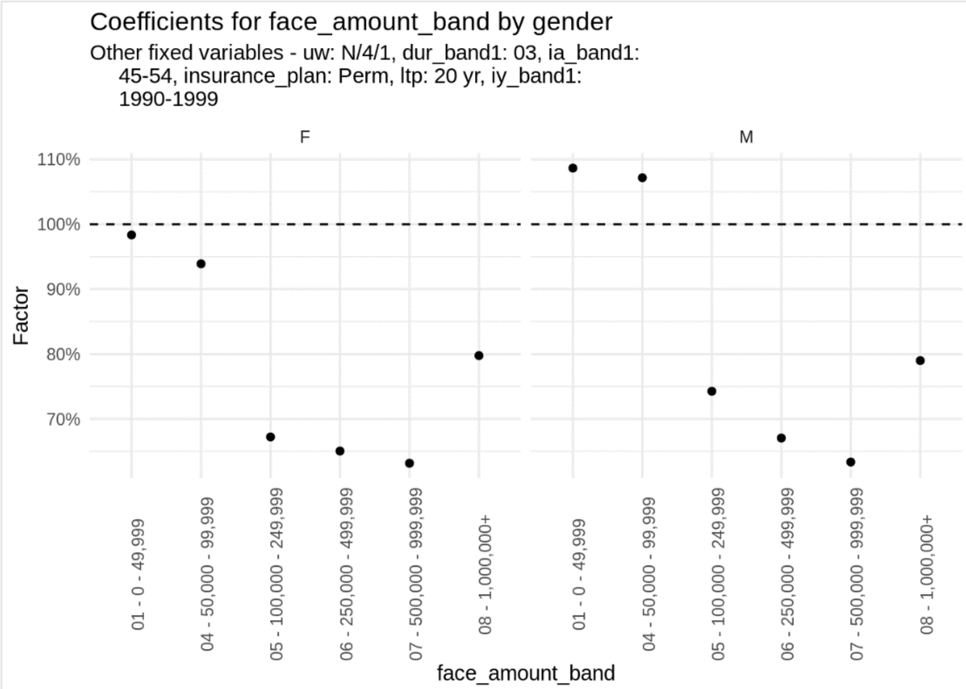


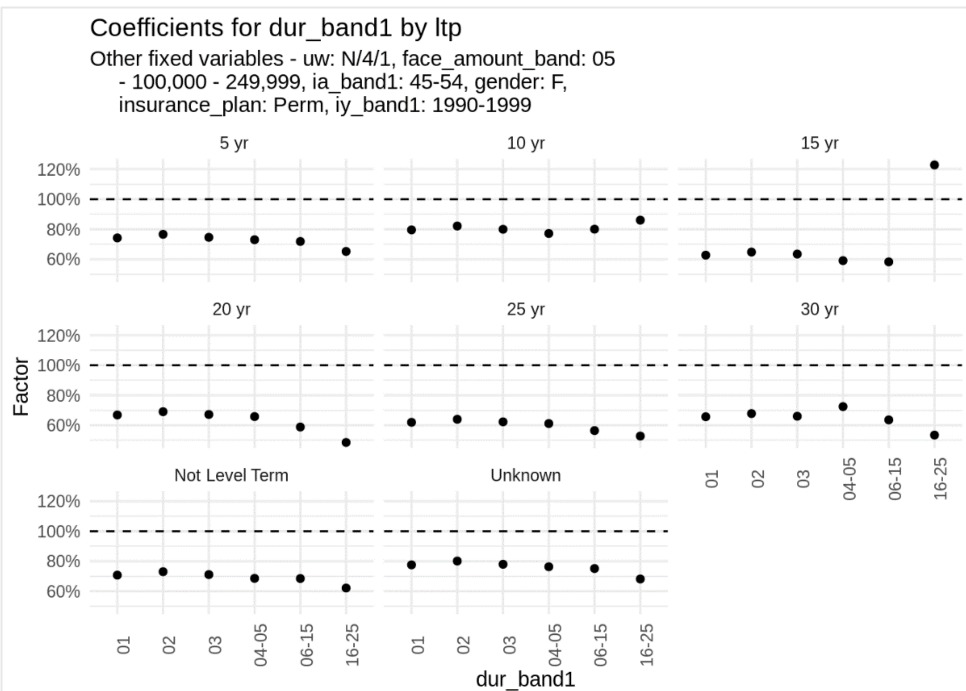

Coefficients for face_amount_band by ia_band1
Other fixed variables - uw: N/4/1, dur_band1: 03, gender: F, insurance_plan: Perm, ltp: 20 yr, iy_band1: 1990-1999
18-24
25-34
35-44
45-54
55-64
65-74
75-84
85-99
120%
100%
80%
60%
40%
Factor
01 - 0 - 49,999
04 - 50,000 - 99,999
05 - 100,000 - 249,999
06 - 250,000 - 499,999
07 - 500,000 - 999,999
08 - 1,000,000+
face_amount_band

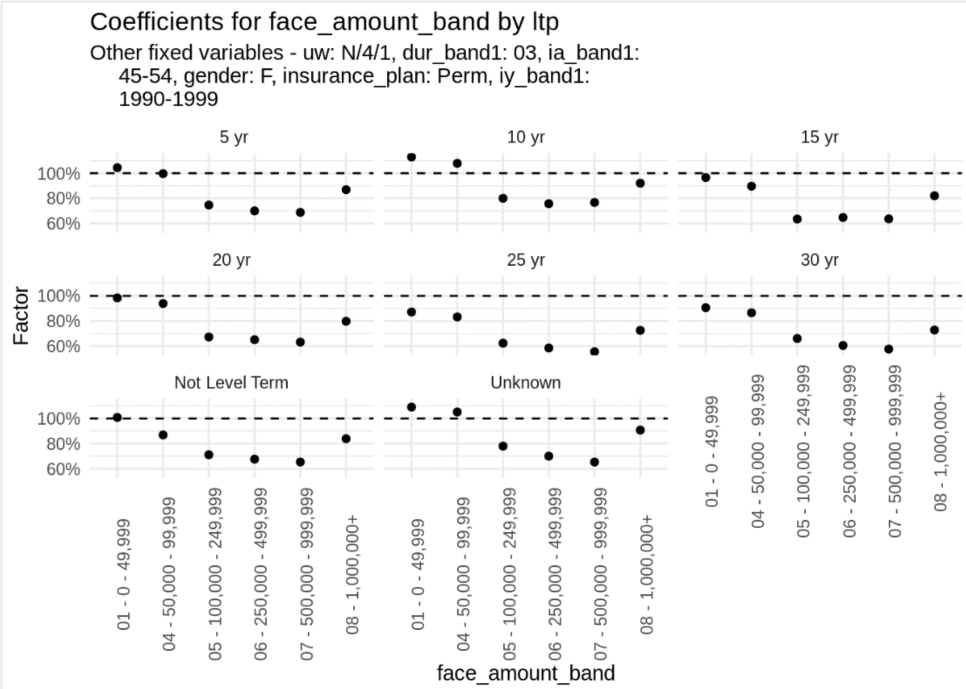

Coefficients for dur_band1 by face_amount_band
Other fixed variables - uw: N/4/1, ia_band1: 45-54, gender: F, insurance_plan: Perm, ltp: 20 yr, iy_band1: 1990-1999
01 - 0 - 49,999
04 - 50,000 - 99,999
05 - 100,000 - 249,999
06 - 250,000 - 499,999
07 - 500,000 - 999,999
08 - 1,000,000+
100%
80%
60%
Factor
01
02
03
04-05
06-15
16-25
dur_band1

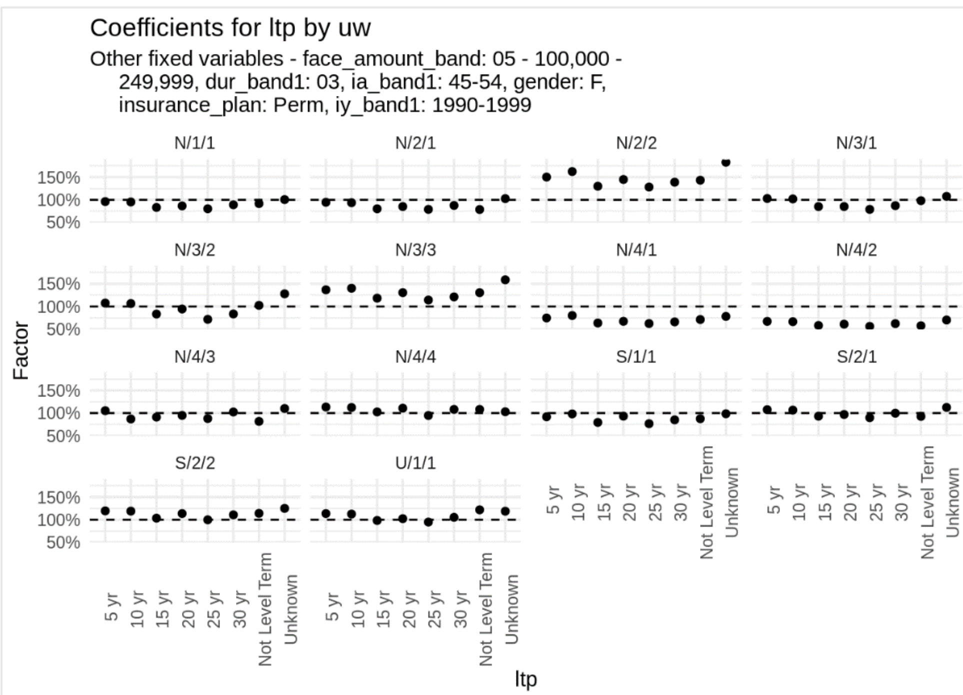

Coefficients for dur_band1 by uw
Other fixed variables - face_amount_band: 05 - 100,000 - 249,999, ia_band1: 45-54, gender: F, insurance_plan: Perm, ltp: 20 yr, iy_band1: 1990-1999
N/1/1
N/2/1
N/2/2
N/3/1
N/3/2
N/3/3
N/4/1
N/4/2
N/4/3
N/4/4
S/1/1
S/2/1
S/2/2
U/1/1
150%
125%
100%
75%
50%
Factor
01
02
03
04-05
06-15
16-25
dur_band1

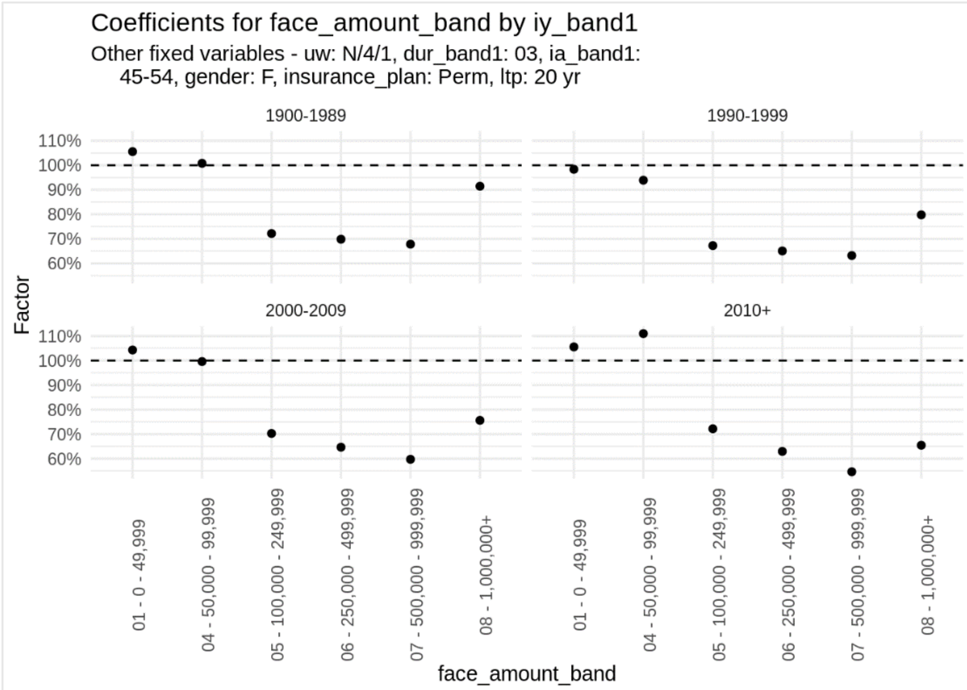

Coefficients for dur_band1 by ia_band1
Other fixed variables - uw: N/4/1, face_amount_band: 05 - 100,000 - 249,999, gender: F, insurance_plan: Perm, ltp: 20 yr, iy_band1: 1990-1999
18-24
25-34
35-44
45-54
55-64
65-74
75-84
85-99
100%
80%
60%
40%
Factor
01
02
03
04-05
06-15
16-25
dur_band1

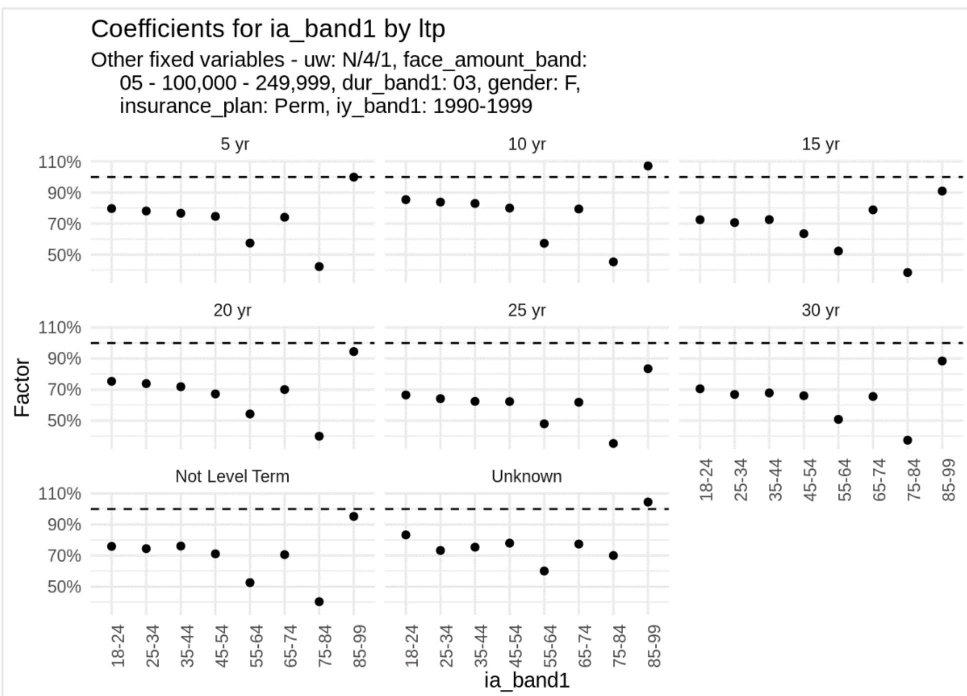

Coefficients for face_amount_band by gender
Other fixed variables - uw: N/4/1, dur_band1: 03, ia_band1: 45-54, insurance_plan: Perm, ltp: 20 yr, iy_band1: 1990-1999
F
M
110%
100%
90%
80%
70%
Factor
01 - 0 - 49,999
04 - 50,000 - 99,999
05 - 100,000 - 249,999
06 - 250,000 - 499,999
07 - 500,000 - 999,999
08 - 1,000,000+
face_amount_band

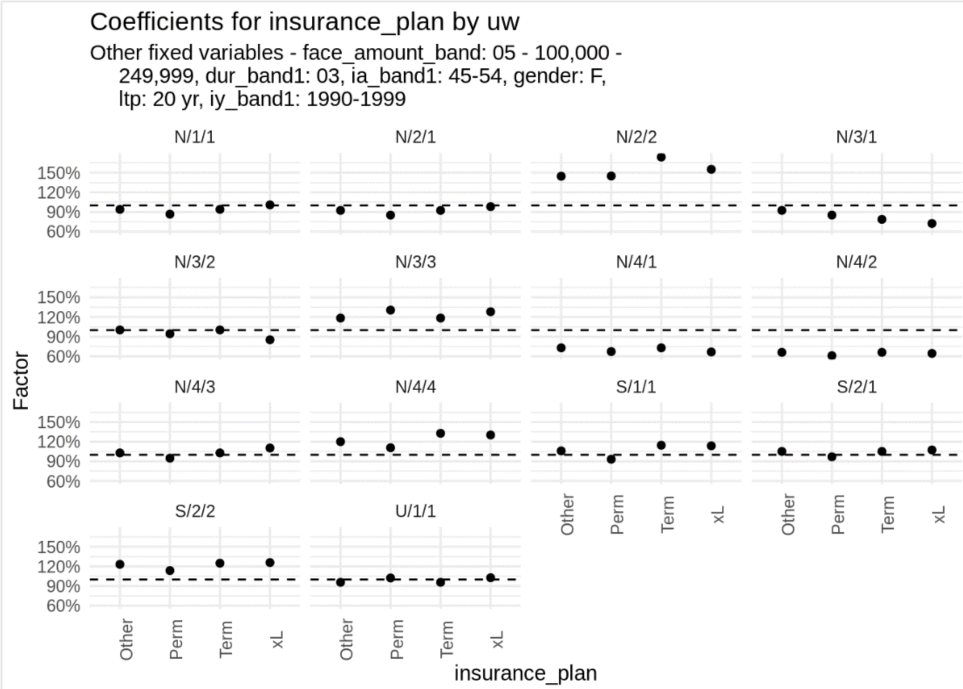

Coefficients for dur_band1 by ltp
Other fixed variables - uw: N/4/1, face_amount_band: 05 - 100,000 - 249,999, ia_band1: 45-54, gender: F, insurance_plan: Perm, iy_band1: 1990-1999
5 yr
10 yr
15 yr
20 yr
25 yr
30 yr
Not Level Term
Unknown
120%
100%
80%
60%
Factor
01
02
03
04-05
06-15
16-25
dur_band1

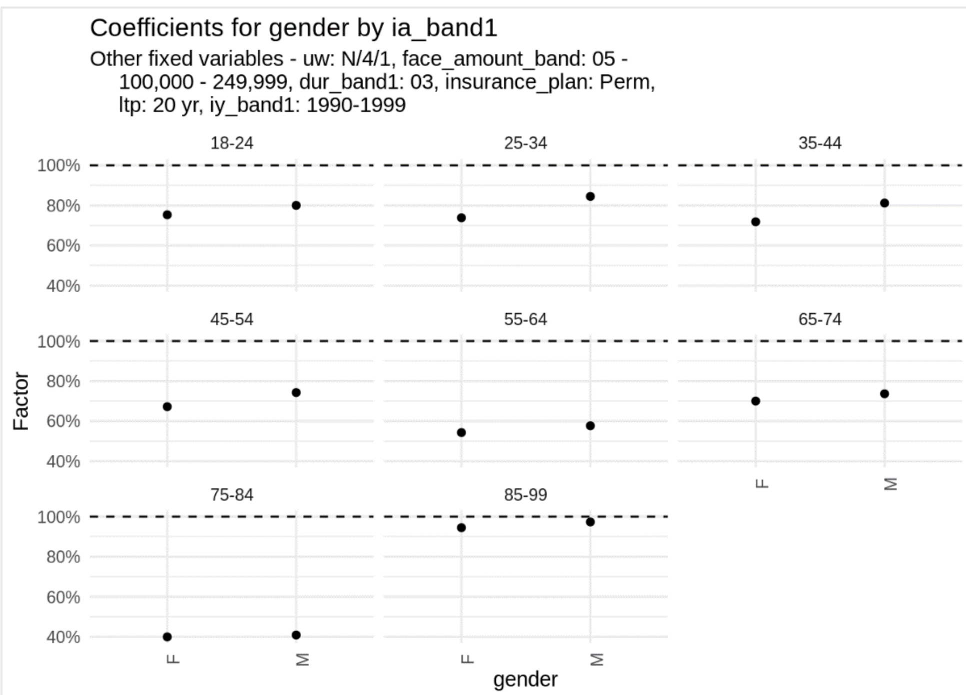

Coefficients for face_amount_band by ltp
Other fixed variables - uw: N/4/1, dur_band1: 03, ia_band1: 45-54, gender: F, insurance_plan: Perm, iy_band1: 1990-1999
5 yr
10 yr
15 yr
20 yr
25 yr
30 yr
Not Level Term
Unknown
Factor
100%
80%
60%
01 - 0 - 49,999
04 - 50,000 - 99,999
05 - 100,000 - 249,999
06 - 250,000 - 499,999
07 - 500,000 - 999,999
08 - 1,000,000+
face_amount_band

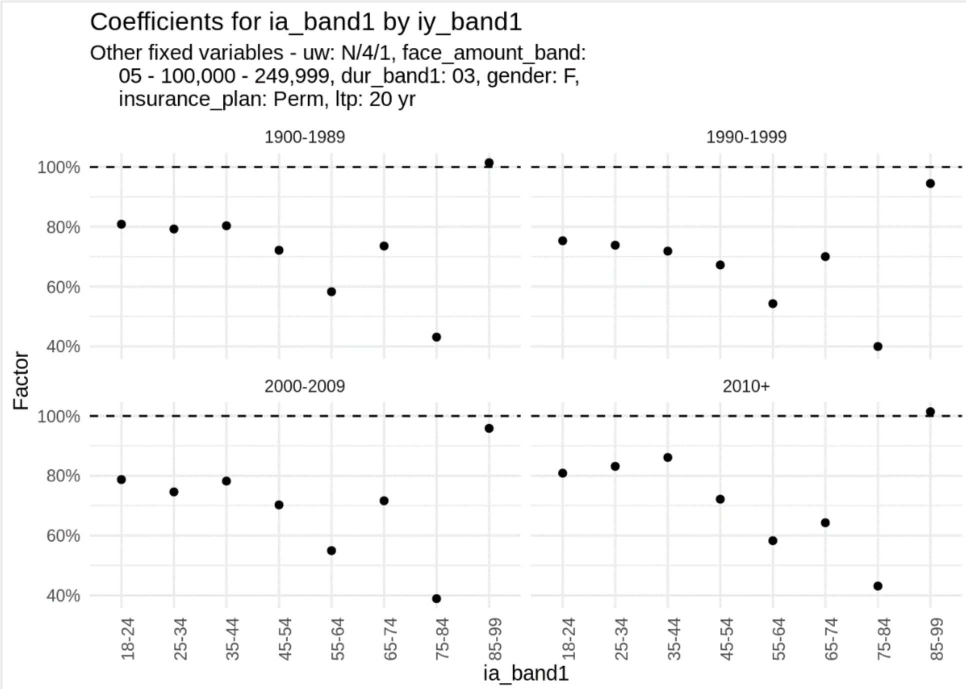

Coefficients for ltp by uw
Other fixed variables - face_amount_band: 05 - 100,000 - 249,999, dur_band1: 03, ia_band1: 45-54, gender: F, insurance_plan: Perm, iy_band1: 1990-1999
N/1/1
N/2/1
N/2/2
N/3/1
N/3/2
N/3/3
N/4/1
N/4/2
N/4/3
N/4/4
S/1/1
S/2/1
S/2/2
U/1/1
Factor
150%
100%
50%
5 yr
10 yr
15 yr
20 yr
25 yr
30 yr
Not Level Term
Unknown
ltp

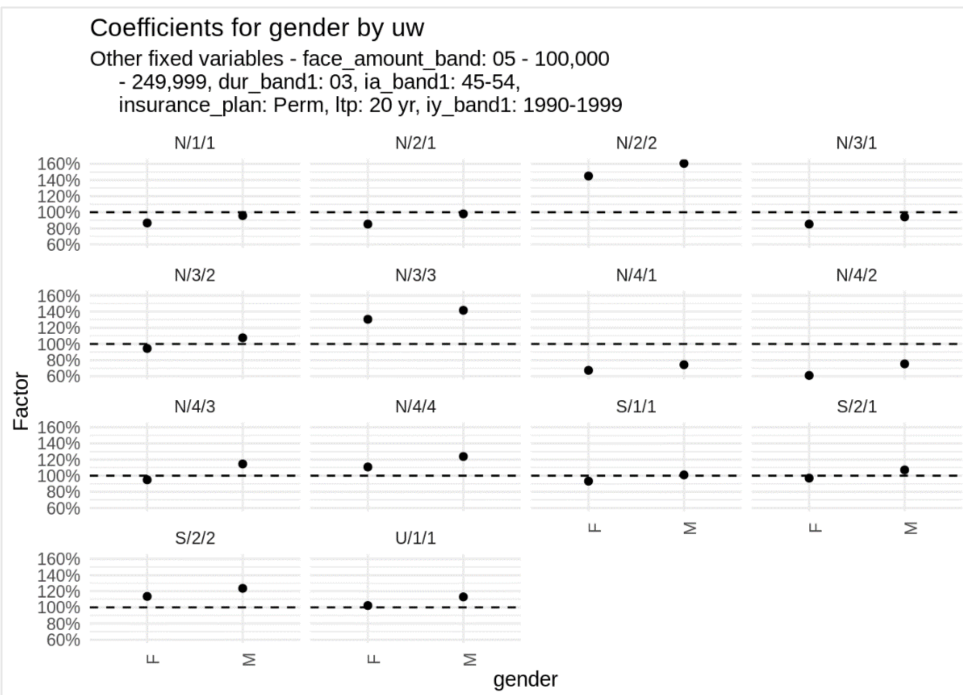

Coefficients for face_amount_band by iy_band1
Other fixed variables - uw: N/4/1, dur_band1: 03, ia_band1: 45-54, gender: F, insurance_plan: Perm, ltp: 20 yr
1900-1989
1990-1999
2000-2009
2010+
Factor
110%
100%
90%
80%
70%
60%
01 - 0 - 49,999
04 - 50,000 - 99,999
05 - 100,000 - 249,999
06 - 250,000 - 499,999
07 - 500,000 - 999,999
08 - 1,000,000+
face_amount_band

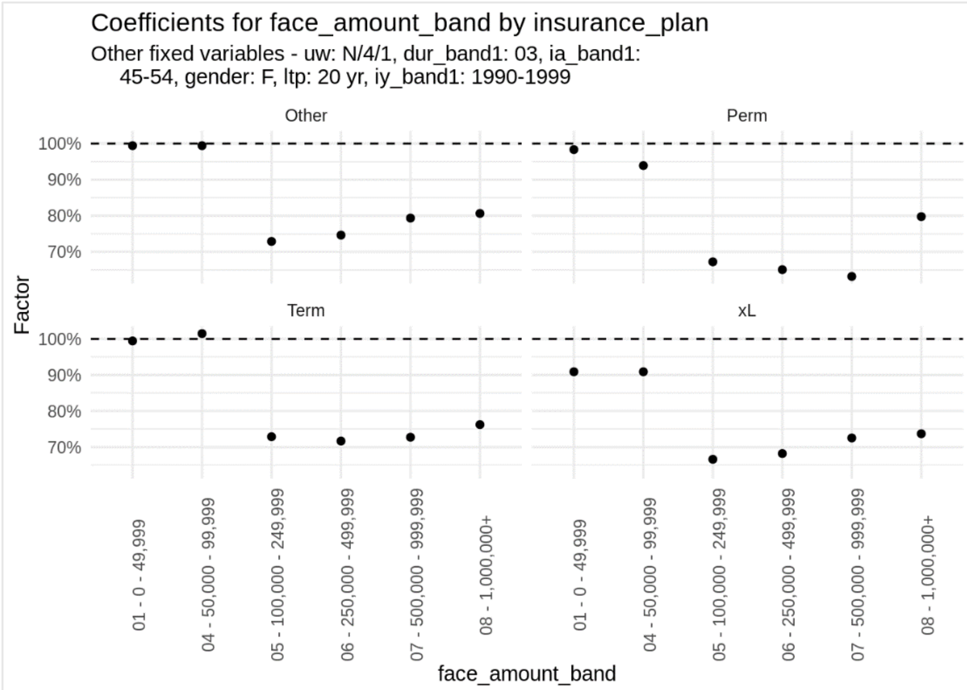

Coefficients for ia_band1 by ltp
Other fixed variables - uw: N/4/1, face_amount_band: 05 - 100,000 - 249,999, dur_band1: 03, gender: F, insurance_plan: Perm, iy_band1: 1990-1999
5 yr
10 yr
15 yr
20 yr
25 yr
30 yr
Not Level Term
Unknown
Factor
110%
90%
70%
50%
18-24
25-34
35-44
45-54
55-64
65-74
75-84
85-99
ia_band1

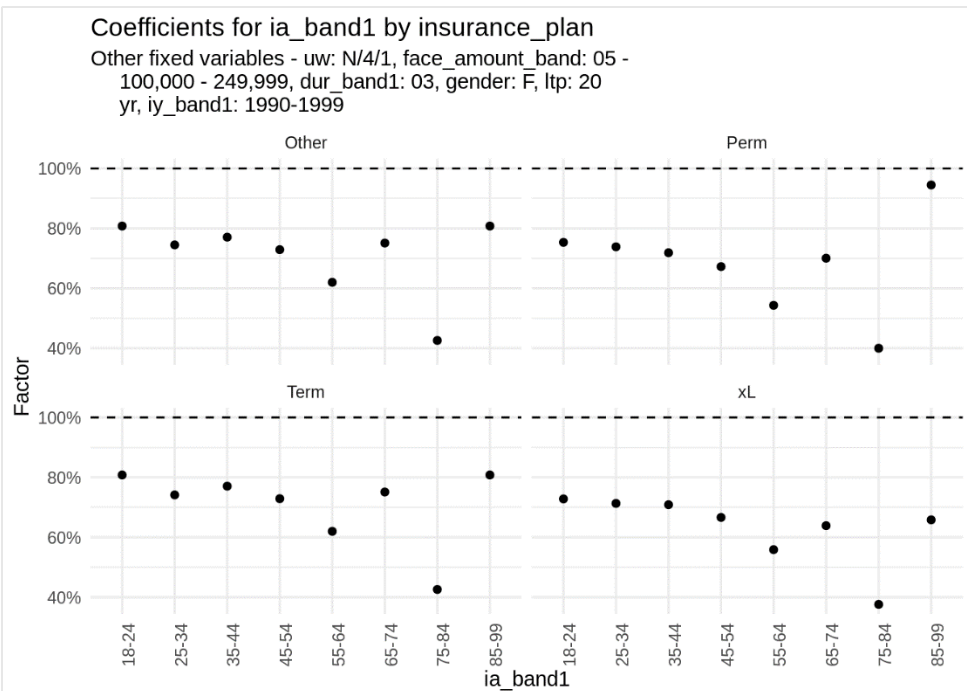

Coefficients for insurance_plan by uw
Other fixed variables - face_amount_band: 05 - 100,000 - 249,999, dur_band1: 03, ia_band1: 45-54, gender: F, ltp: 20 yr, iy_band1: 1990-1999
N/1/1
N/2/1
N/2/2
N/3/1
N/3/2
N/3/3
N/4/1
N/4/2
N/4/3
N/4/4
S/1/1
S/2/1
S/2/2
U/1/1
150%
120%
90%
60%
Factor
Other
Perm
Term
xL
insurance_plan

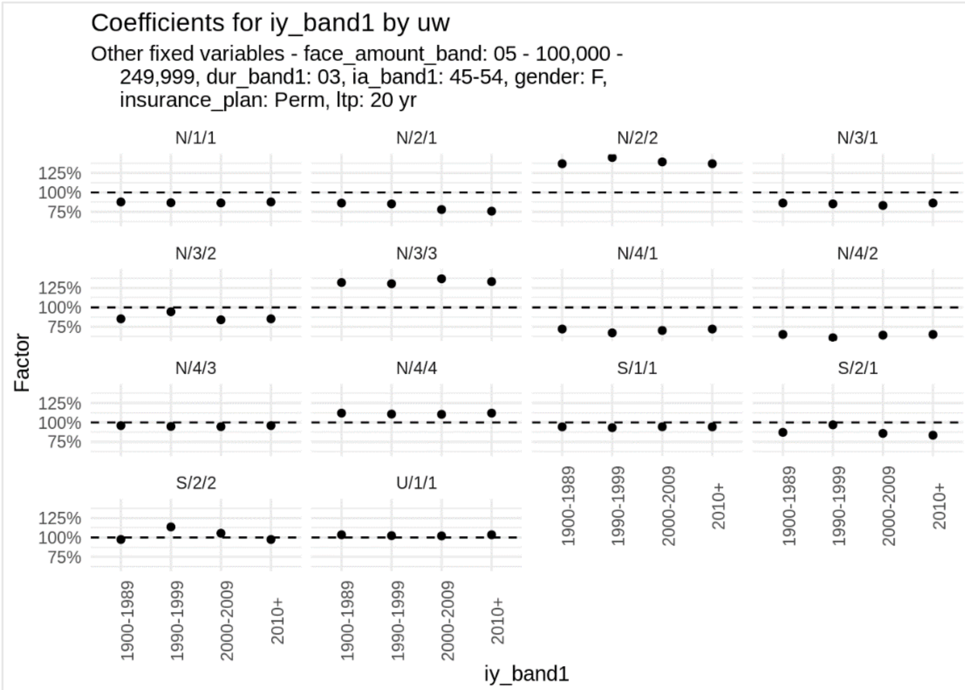

Coefficients for gender by ia_band1
Other fixed variables - uw: N/4/1, face_amount_band: 05 - 100,000 - 249,999, dur_band1: 03, insurance_plan: Perm, ltp: 20 yr, iy_band1: 1990-1999
18-24
25-34
35-44
45-54
55-64
65-74
75-84
85-99
100%
80%
60%
40%
Factor
F
M
gender

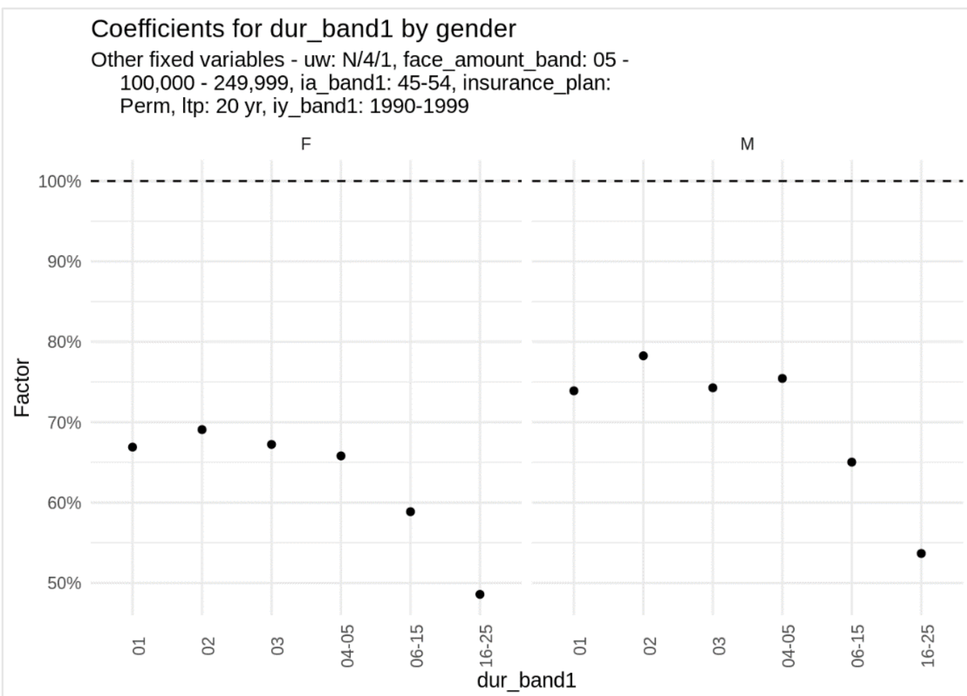

Coefficients for ia_band1 by iy_band1
Other fixed variables - uw: N/4/1, face_amount_band: 05 - 100,000 - 249,999, dur_band1: 03, gender: F, insurance_plan: Perm, ltp: 20 yr
1900-1989
1990-1999
2000-2009
2010+
Factor
100%
80%
60%
40%
18-24
25-34
35-44
45-54
55-64
65-74
75-84
85-99
ia_band1

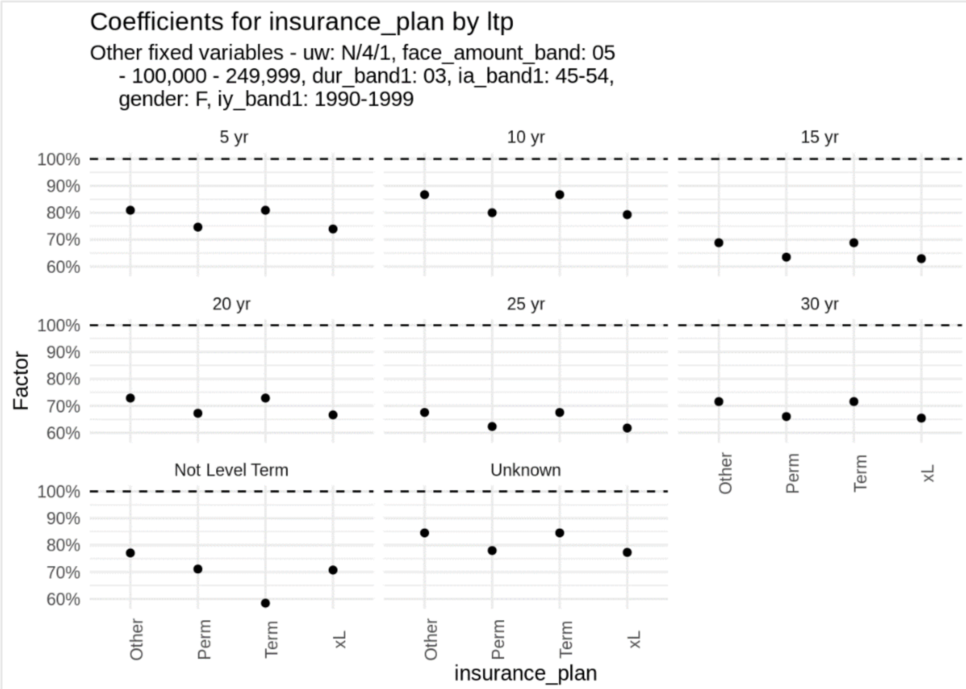

Coefficients for gender by uw
Other fixed variables - face_amount_band: 05 - 100,000 - 249,999, dur_band1: 03, ia_band1: 45-54, insurance_plan: Perm, ltp: 20 yr, iy_band1: 1990-1999
N/1/1
N/2/1
N/2/2
N/3/1
N/3/2
N/3/3
N/4/1
N/4/2
N/4/3
N/4/4
S/1/1
S/2/1
S/2/2
U/1/1
Factor
160%
140%
120%
100%
80%
60%
F
M
gender

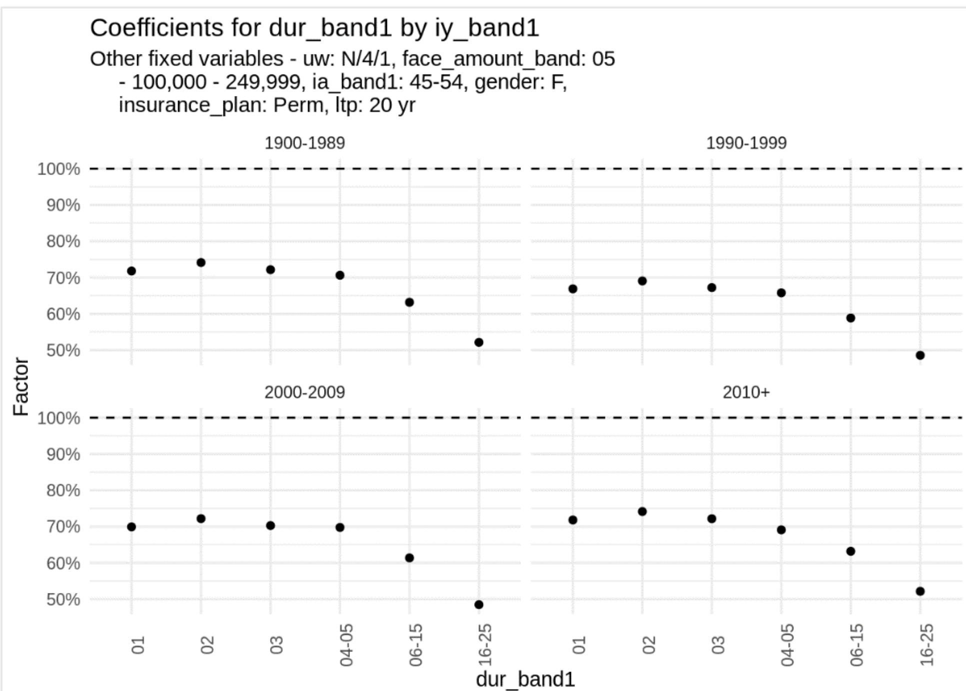

Coefficients for face_amount_band by insurance_plan
Other fixed variables - uw: N/4/1, dur_band1: 03, ia_band1: 45-54, gender: F, ltp: 20 yr, iy_band1: 1990-1999
Other
Perm
Term
xL
Factor
100%
90%
80%
70%
01 - 0 - 49,999
04 - 50,000 - 99,999
05 - 100,000 - 249,999
06 - 250,000 - 499,999
07 - 500,000 - 999,999
08 - 1,000,000+
face_amount_band

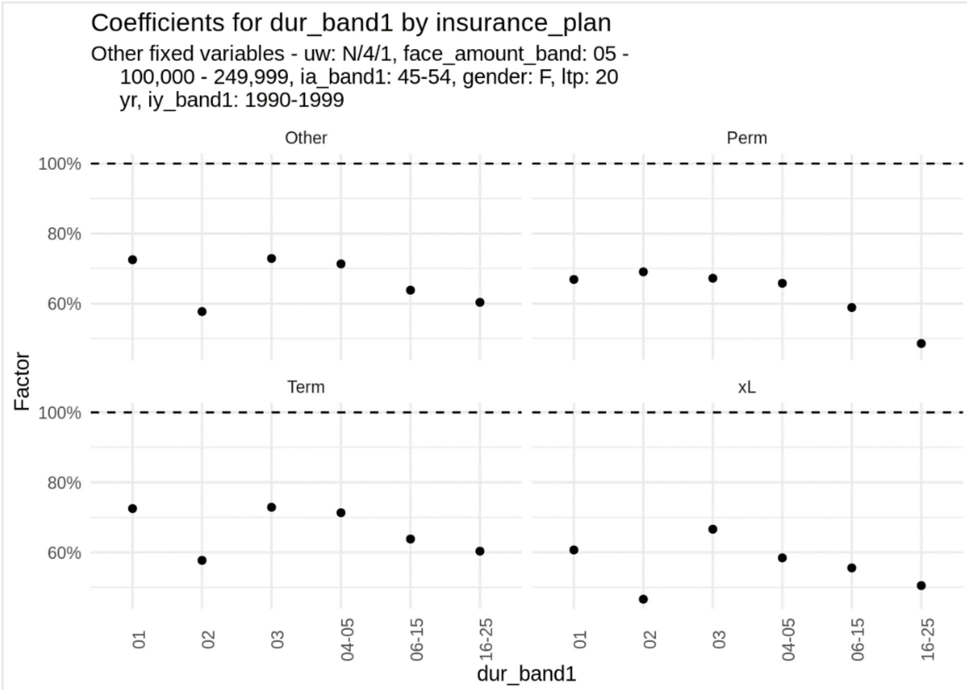

Coefficients for ia_band1 by insurance_plan
Other fixed variables - uw: N/4/1, face_amount_band: 05 - 100,000 - 249,999, dur_band1: 03, gender: F, ltp: 20 yr, iy_band1: 1990-1999
Other
Perm
Term
xL
Factor
100%
80%
60%
40%
18-24
25-34
35-44
45-54
55-64
65-74
75-84
85-99
ia_band1

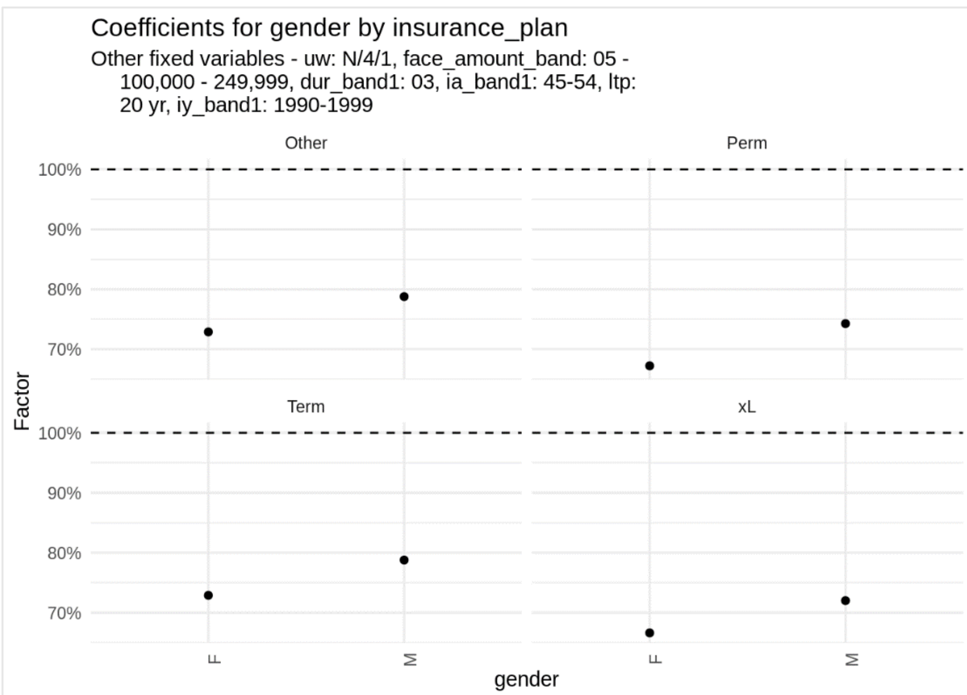

Coefficients for iy_band1 by uw
Other fixed variables - face_amount_band: 05 - 100,000 - 249,999, dur_band1: 03, ia_band1: 45-54, gender: F, insurance_plan: Perm, ltp: 20 yr
N/1/1
N/2/1
N/2/2
N/3/1
N/3/2
N/3/3
N/4/1
N/4/2
N/4/3
N/4/4
S/1/1
S/2/1
S/2/2
U/1/1
Factor
125%
100%
75%
1900-1989
1990-1999
2000-2009
2010+
iy_band1

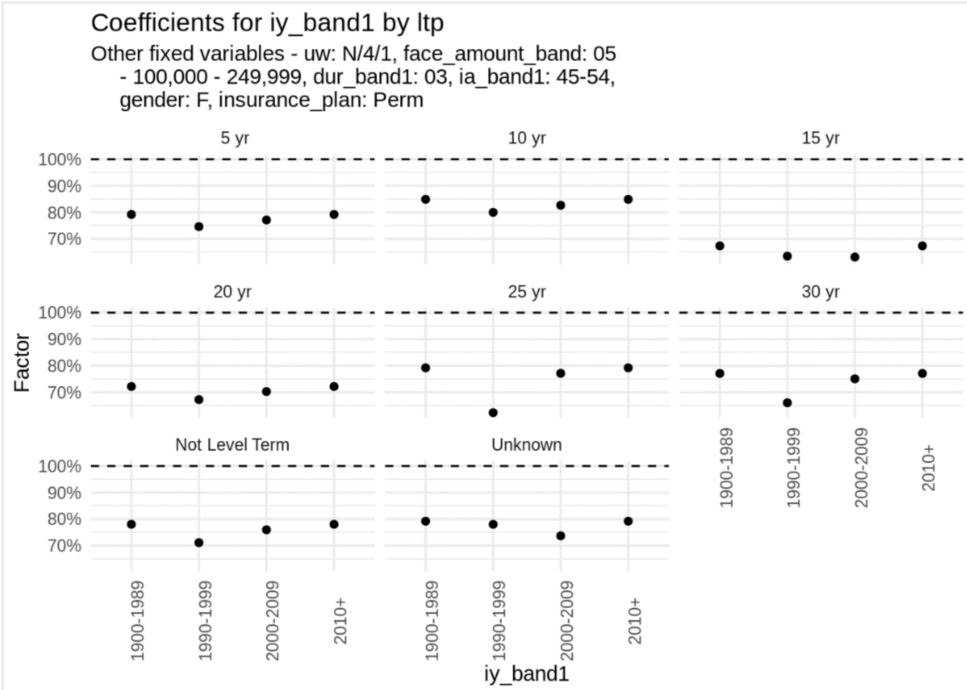

Coefficients for dur_band1 by gender
Other fixed variables - uw: N/4/1, face_amount_band: 05 - 100,000 - 249,999, ia_band1: 45-54, insurance_plan: Perm, ltp: 20 yr, iy_band1: 1990-1999
F
M
Factor
100%
90%
80%
70%
60%
50%
01
02
03
04-05
06-15
16-25
dur_band1

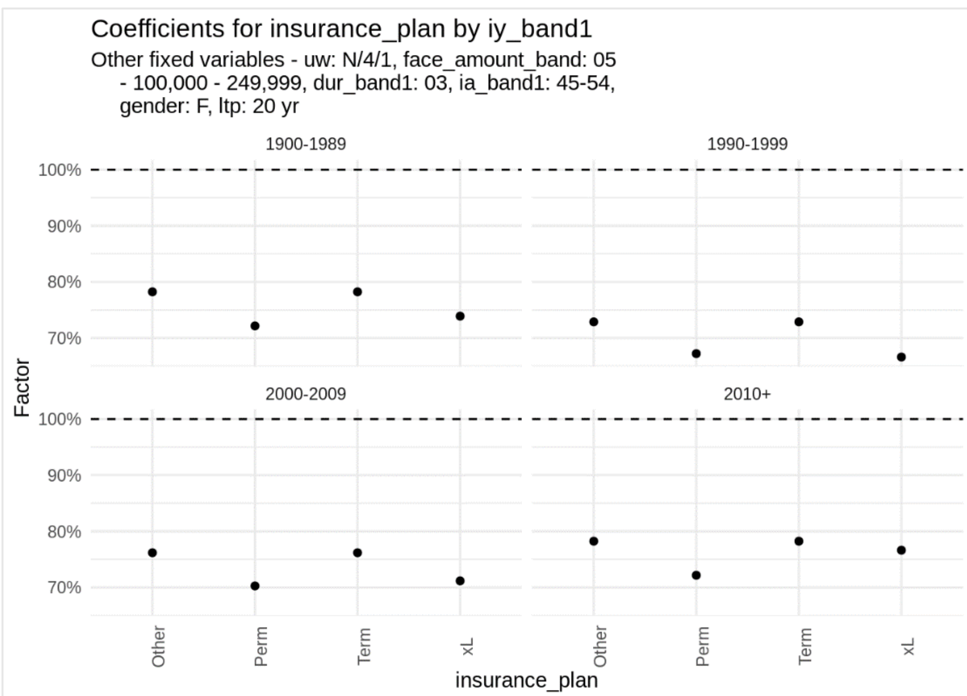

Coefficients for insurance_plan by ltp
Other fixed variables - uw: N/4/1, face_amount_band: 05 - 100,000 - 249,999, dur_band1: 03, ia_band1: 45-54, gender: F, iy_band1: 1990-1999
5 yr
10 yr
15 yr
20 yr
25 yr
30 yr
Not Level Term
Unknown
100%
90%
80%
70%
60%
Factor
Other
Perm
Term
xL
insurance_plan

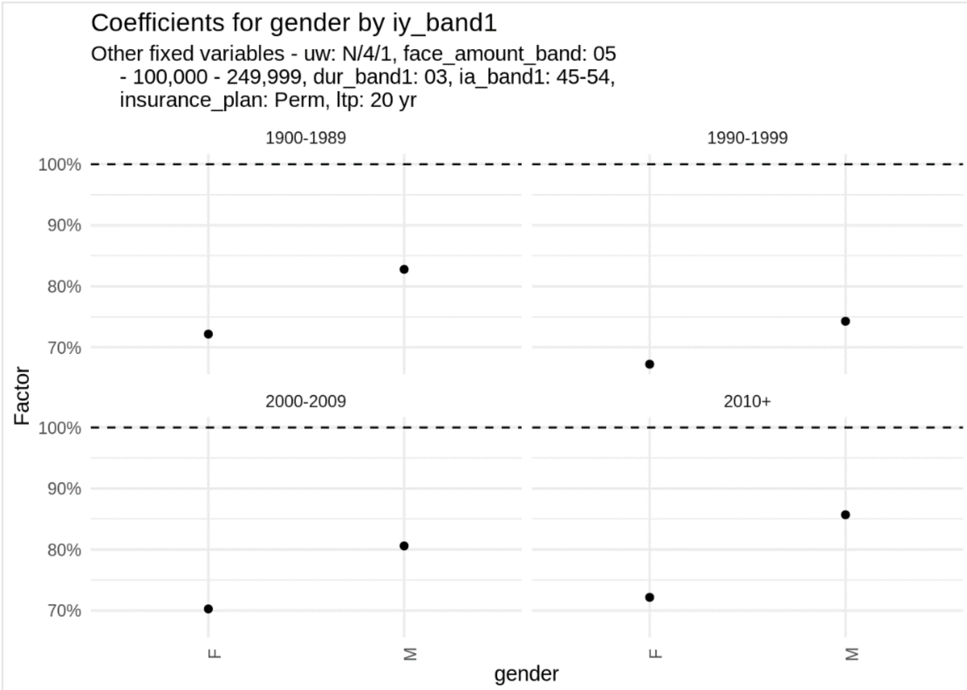

Coefficients for dur_band1 by iy_band1
Other fixed variables - uw: N/4/1, face_amount_band: 05 - 100,000 - 249,999, ia_band1: 45-54, gender: F, insurance_plan: Perm, ltp: 20 yr
1900-1989
1990-1999
2000-2009
2010+
100%
90%
80%
70%
60%
50%
Factor
01
02
03
04-05
06-15
16-25
dur_band1

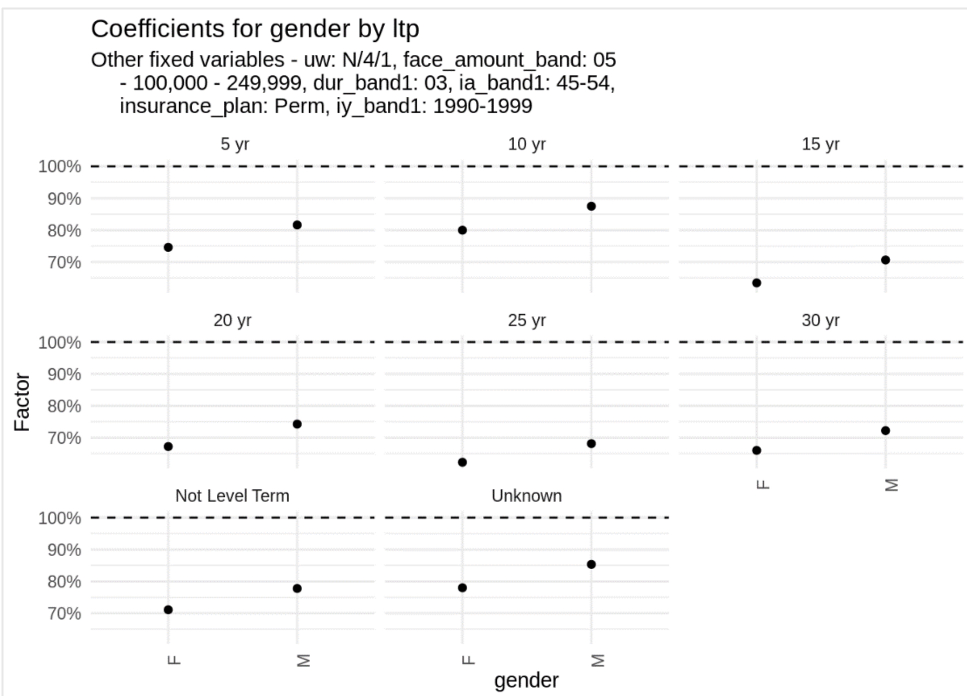

Coefficients for dur_band1 by insurance_plan
Other fixed variables - uw: N/4/1, face_amount_band: 05 - 100,000 - 249,999, ia_band1: 45-54, gender: F, ltp: 20 yr, iy_band1: 1990-1999
Other
Perm
Term
xL
100%
80%
60%
Factor
01
02
03
04-05
06-15
16-25
dur_band1


Coefficients for gender by insurance_plan
Other fixed variables - uw: N/4/1, face_amount_band: 05 - 100,000 - 249,999, dur_band1: 03, ia_band1: 45-54, ltp: 20 yr, iy_band1: 1990-1999
Other
Perm
Term
xL
100%
90%
80%
70%
Factor
F
M
gender


Coefficients for iy_band1 by ltp
Other fixed variables - uw: N/4/1, face_amount_band: 05 - 100,000 - 249,999, dur_band1: 03, ia_band1: 45-54, gender: F, insurance_plan: Perm
5 yr
10 yr
15 yr
20 yr
25 yr
30 yr
Not Level Term
Unknown
100%
90%
80%
70%
Factor
1900-1989
1990-1999
2000-2009
2010+
iy_band1


Coefficients for insurance_plan by iy_band1
Other fixed variables - uw: N/4/1, face_amount_band: 05 - 100,000 - 249,999, dur_band1: 03, ia_band1: 45-54, gender: F, ltp: 20 yr
1900-1989
1990-1999
2000-2009
2010+
100%
90%
80%
70%
Factor
Other
Perm
Term
xL
insurance_plan


Coefficients for gender by iy_band1
Other fixed variables - uw: N/4/1, face_amount_band: 05 - 100,000 - 249,999, dur_band1: 03, ia_band1: 45-54, insurance_plan: Perm, ltp: 20 yr
1900-1989
1990-1999
2000-2009
2010+
100%
90%
80%
70%
Factor
F
M
gender

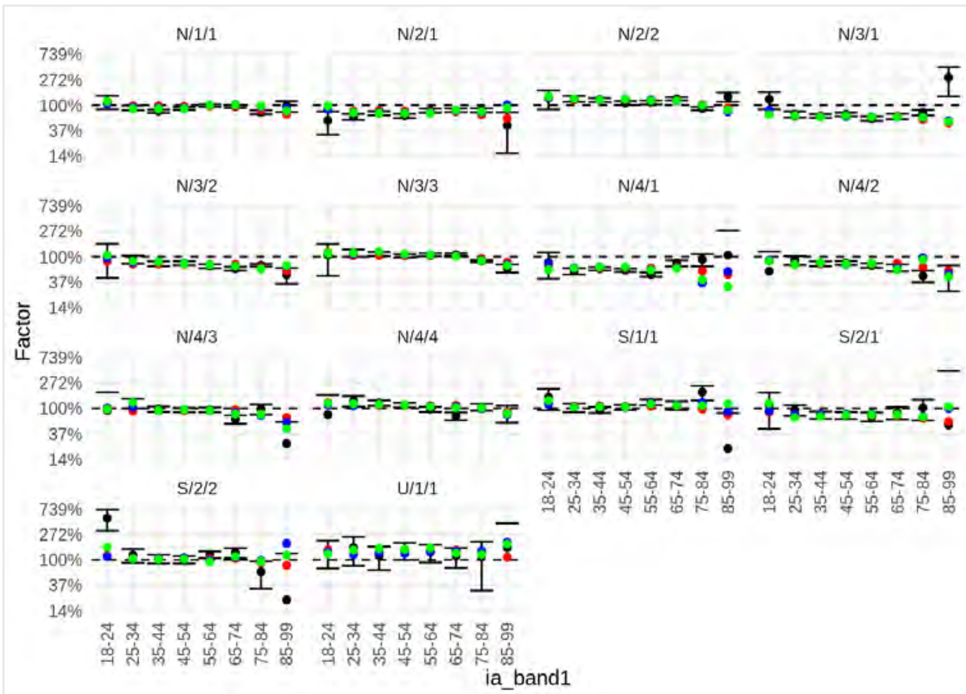

Coefficients for gender by ltp
Other fixed variables - uw: N/4/1, face_amount_band: 05 - 100,000 - 249,999, dur_band1: 03, ia_band1: 45-54, insurance_plan: Perm, iy_band1: 1990-1999
5 yr
10 yr
15 yr
20 yr
25 yr
30 yr
Not Level Term
Unknown
100%
90%
80%
70%
Factor
F
M
gender

APPENDIX C – GRAPHICAL COMPARISON OF MODELS

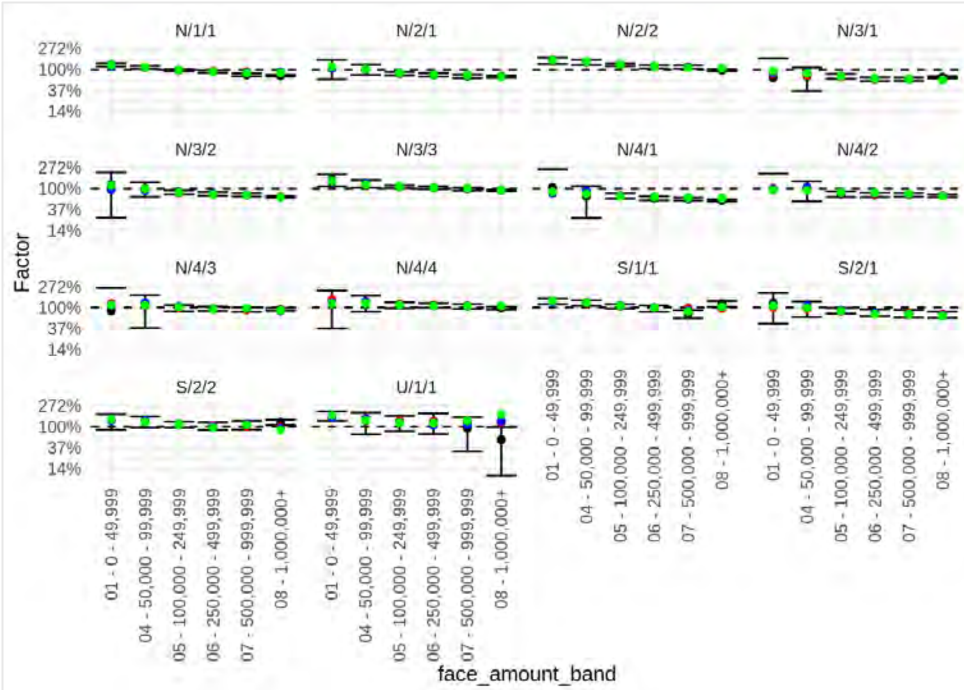
Below are plots of how the model performs versus marginal effects, with performance tested on the test subset. Black dots with error bars are from the actual-to-2015VBT ratio, with error bar width based on the dispersion from the GLM model. (Caution: this is at best a crude approximation.)

The following colors denote specific predictive model ratios versus the 2015 VBT:

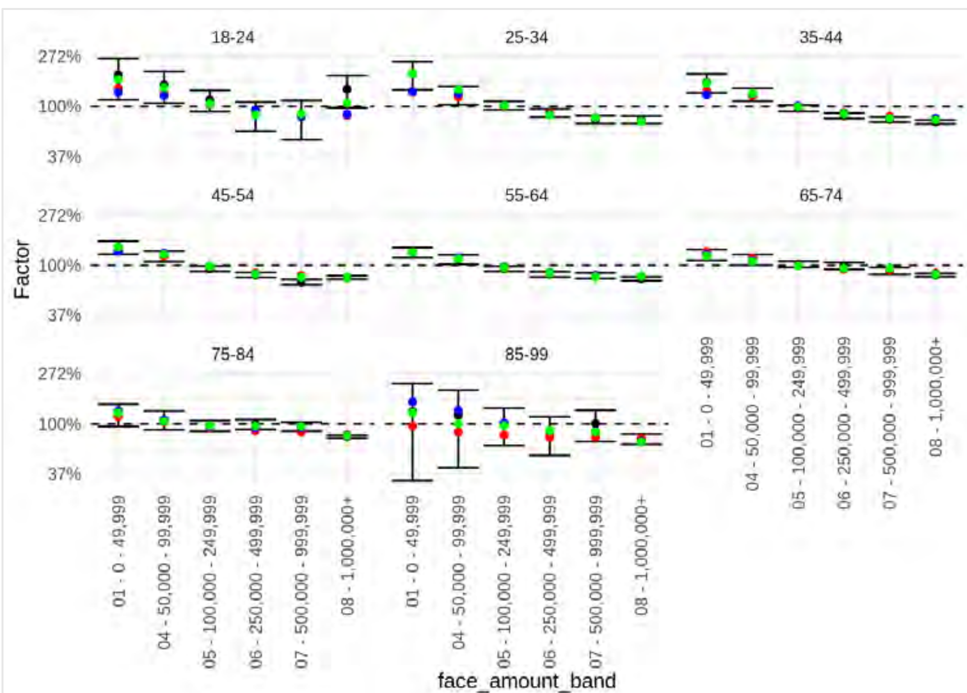
- Red uses GLM predicted claims
- Blue uses predicted claims from the elastic net GLM
- Green uses LightGBM predicted claims



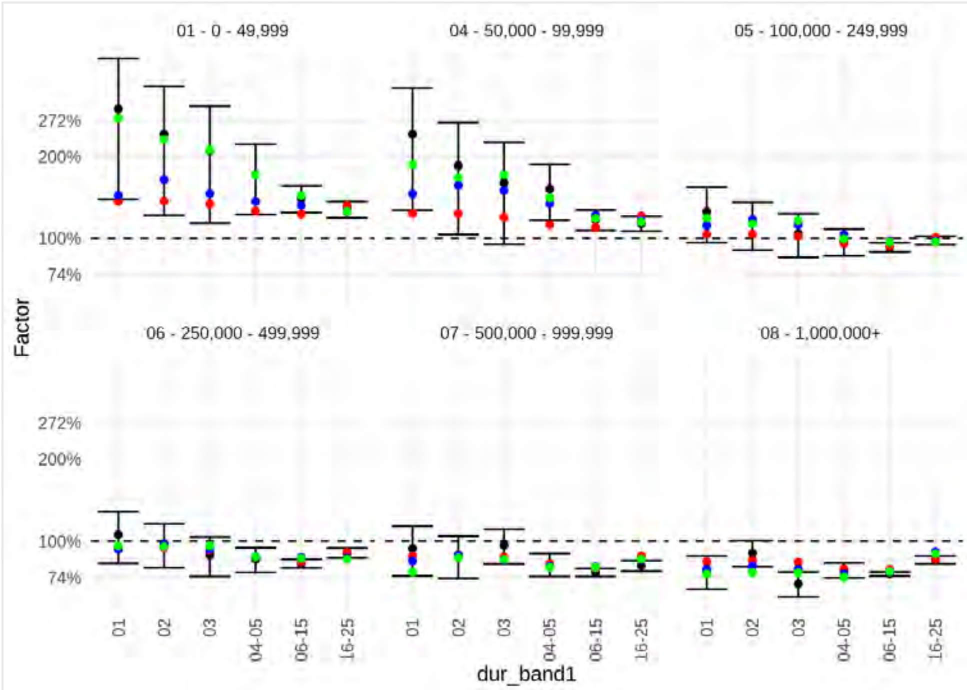
1ia_band1 by uw



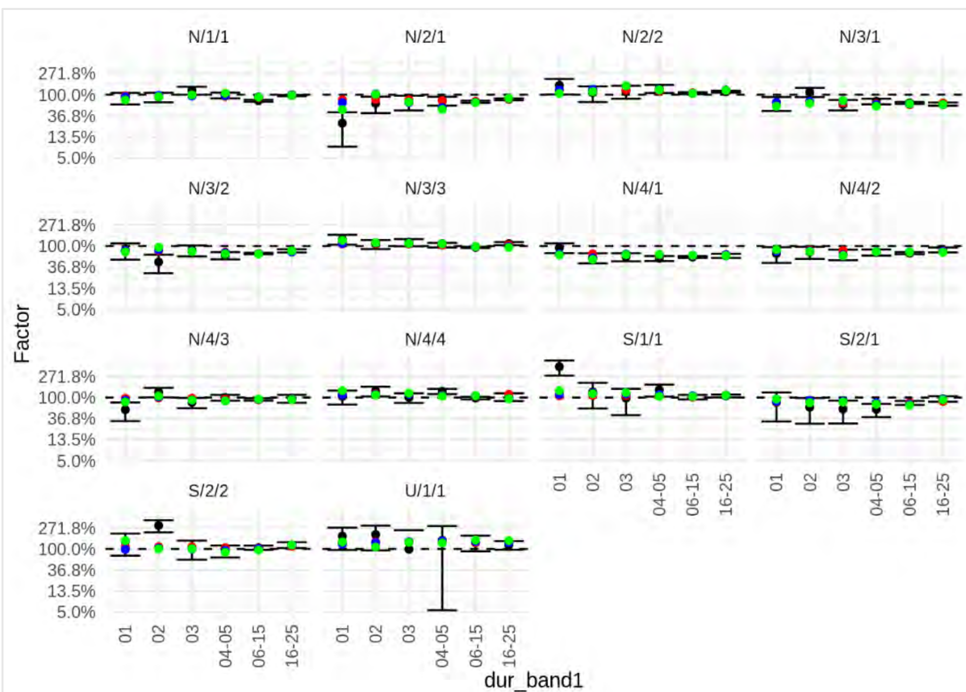
2face_amount_band by uw



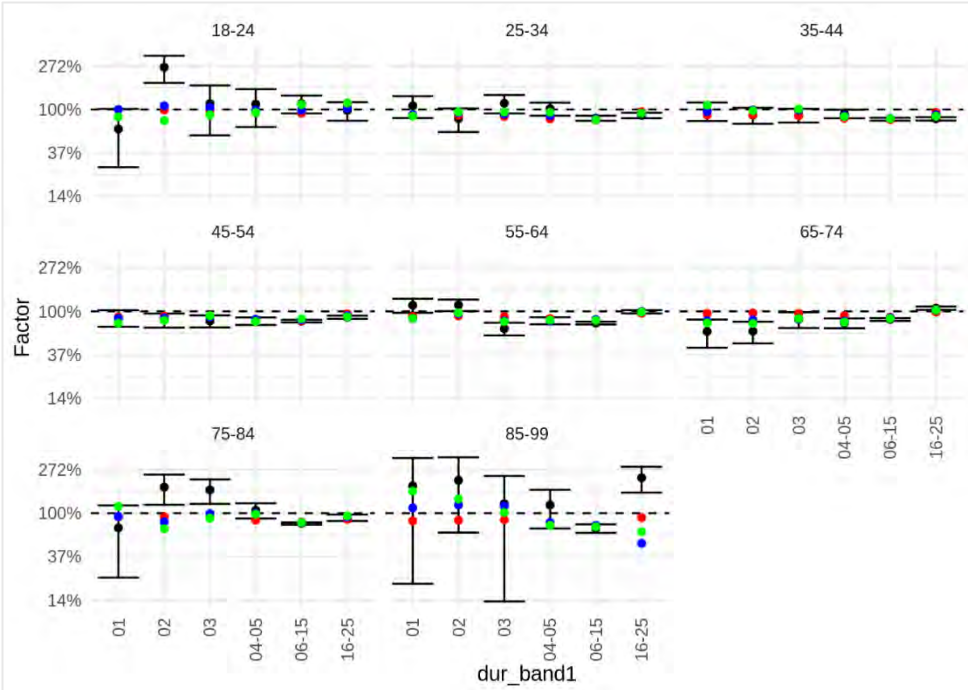
3face_amount_band by ia_band1



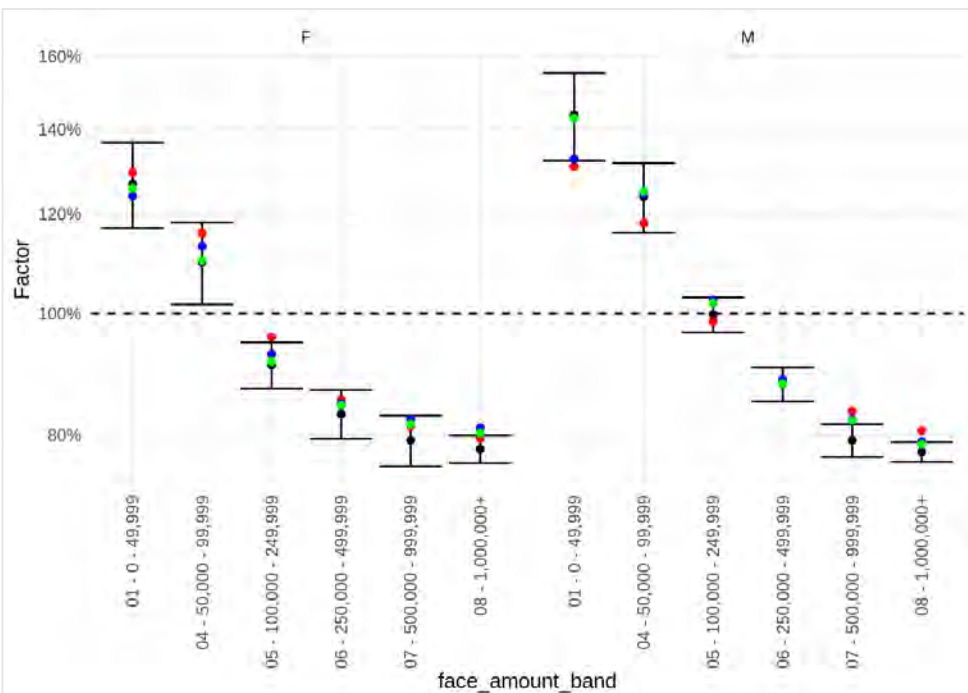
4dur_band1 by face_amount_band



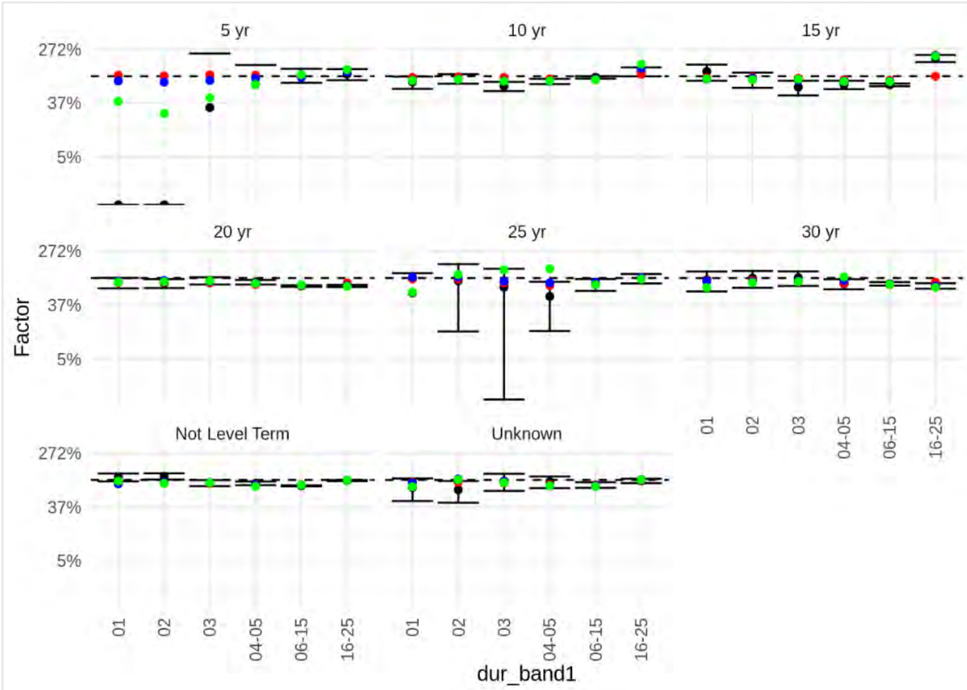
5dur_band1 by uw



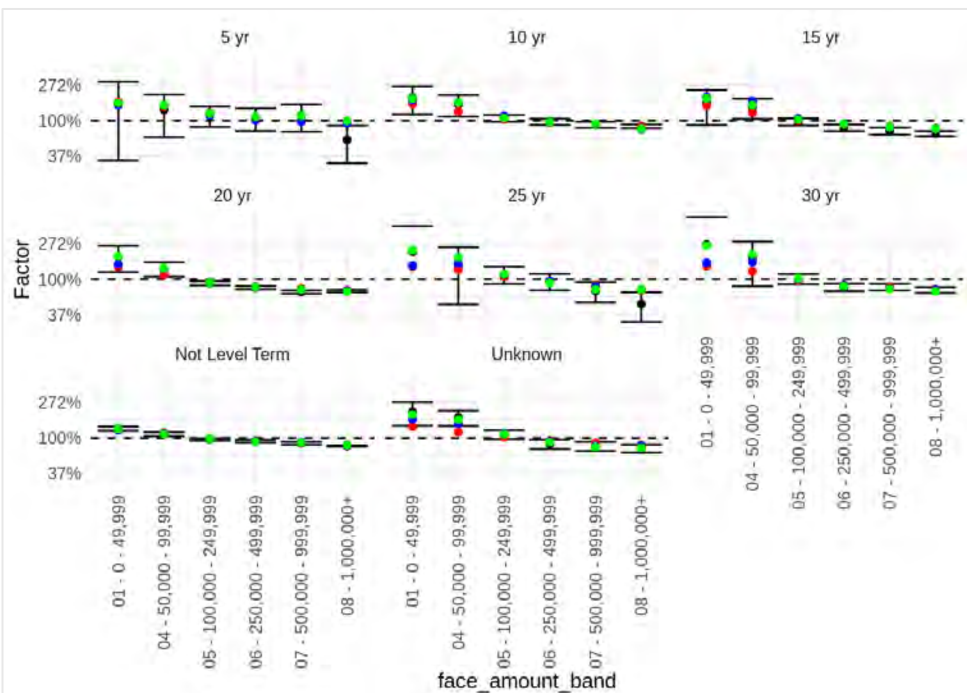
6dur_band1 by ia_band1



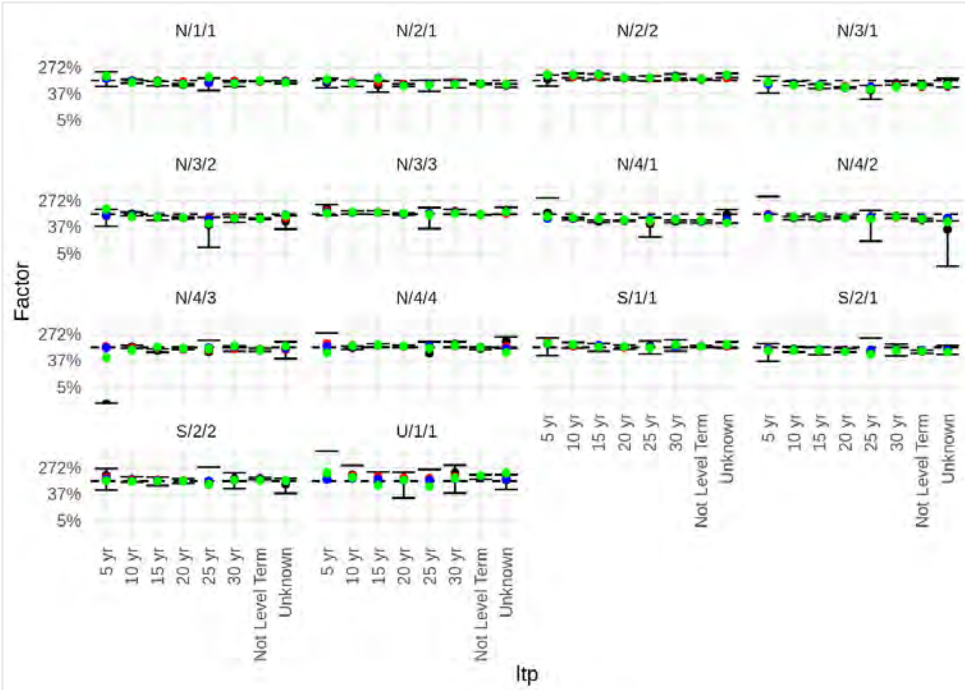
7face_amount_band by gender



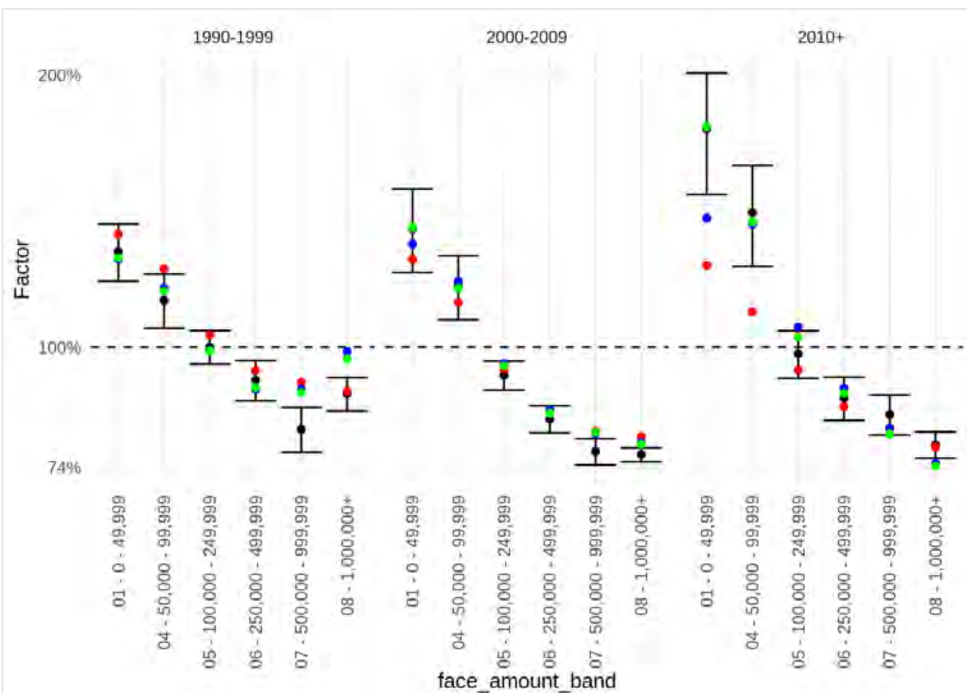
8dur_band1 by ltp



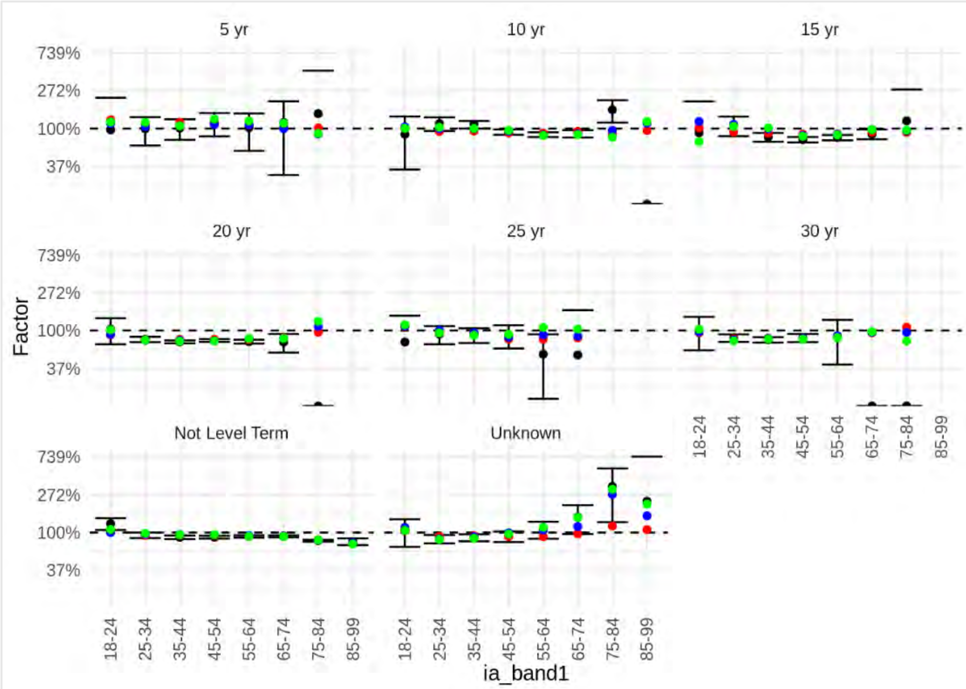
9face_amount_band by ltp



10ltp by uw



11face_amount_band by iy_band1



12ia_band1 by ltp

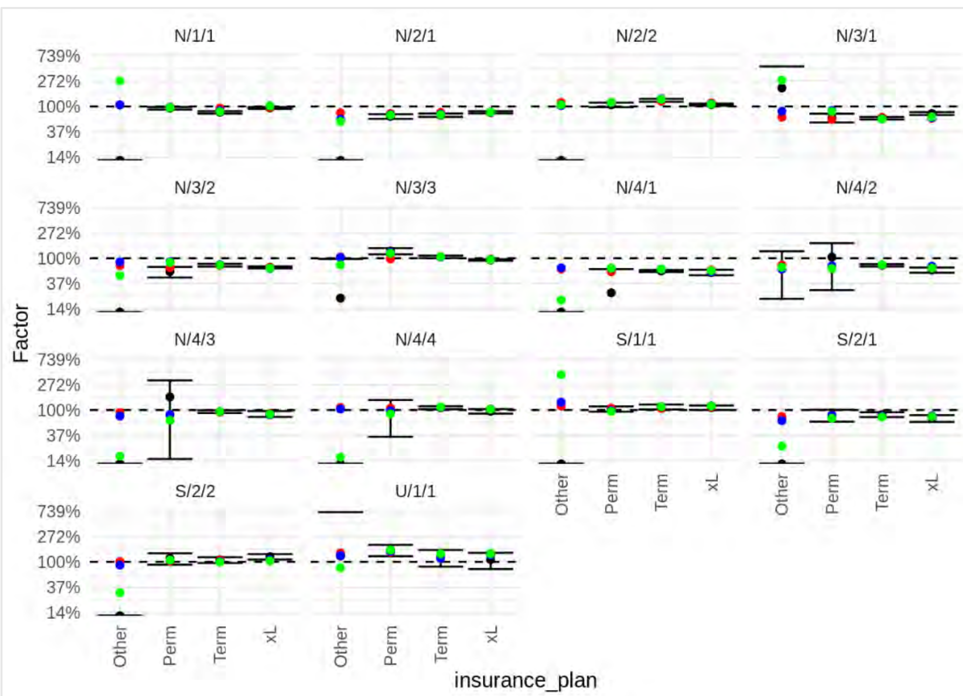
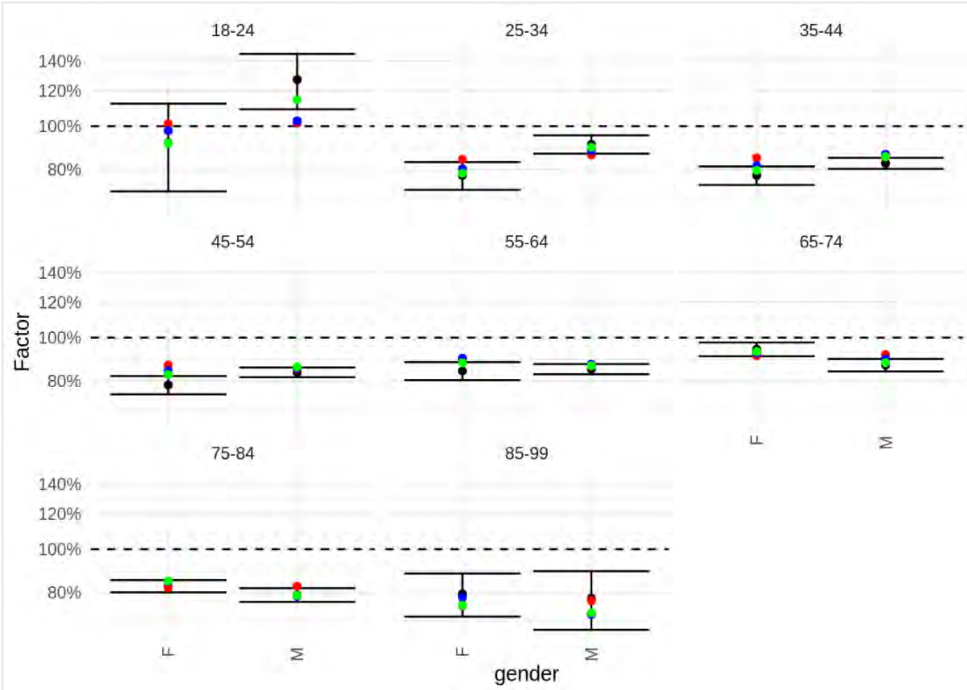
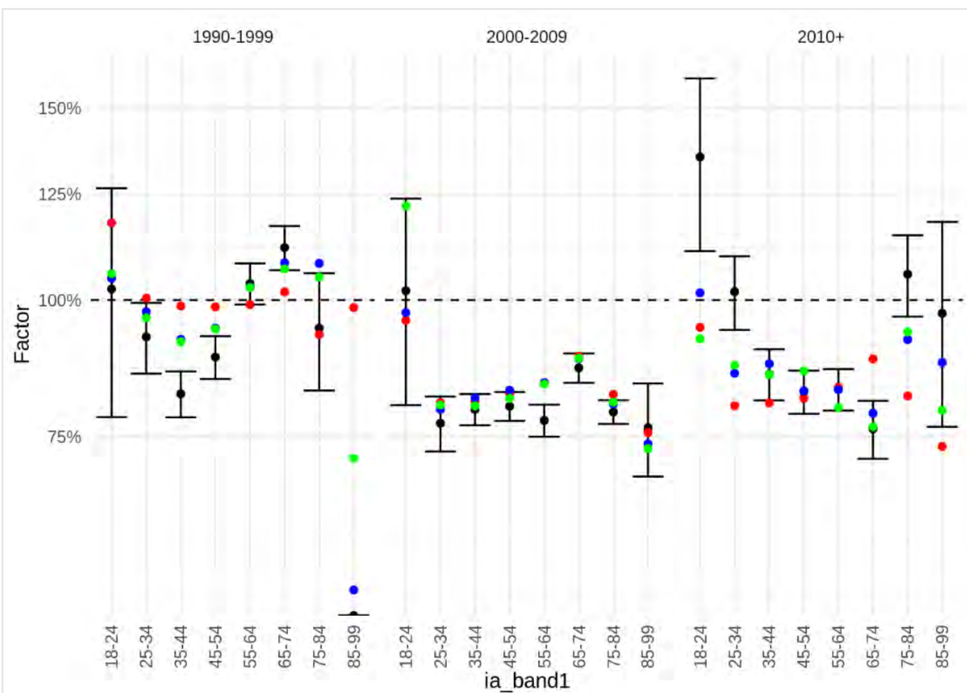


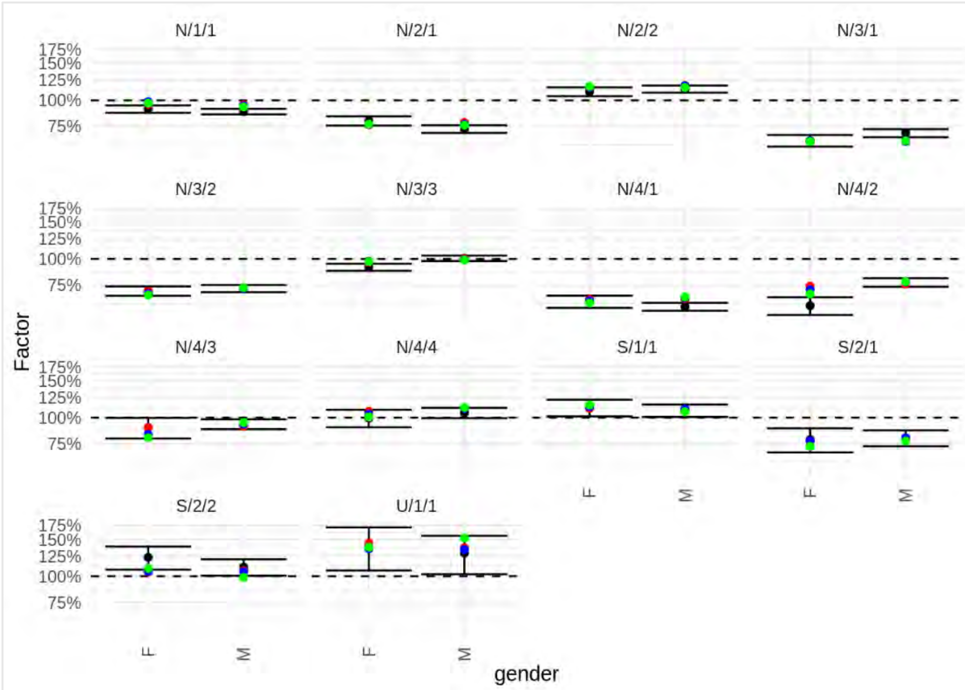
Figure 13 - insurance_plan by uw



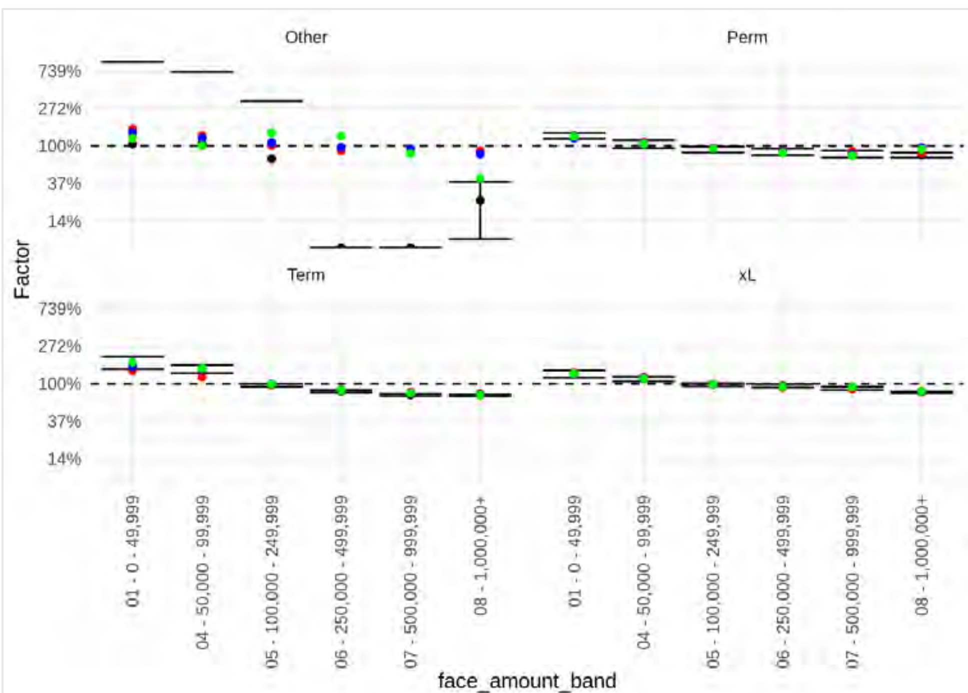
14gender by ia_band1



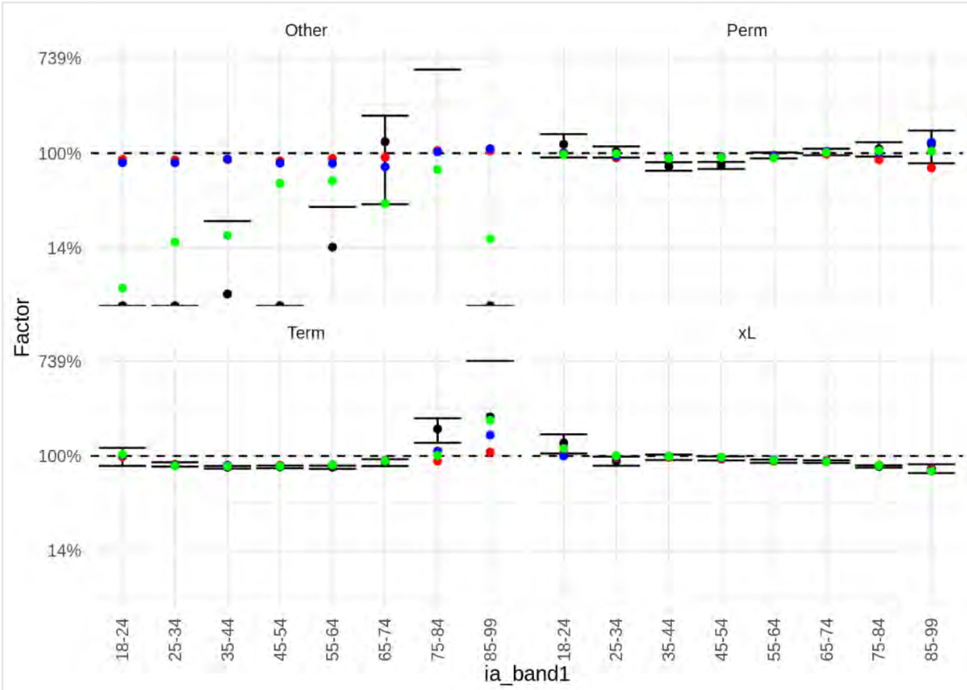
15ia_band1 by iy_band1



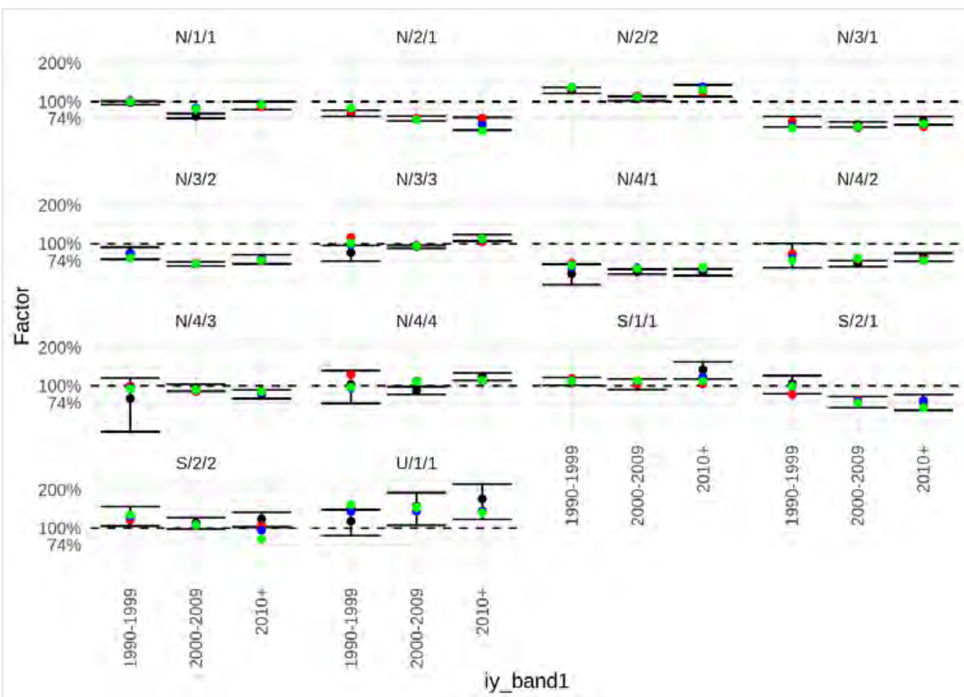
16gender by uw



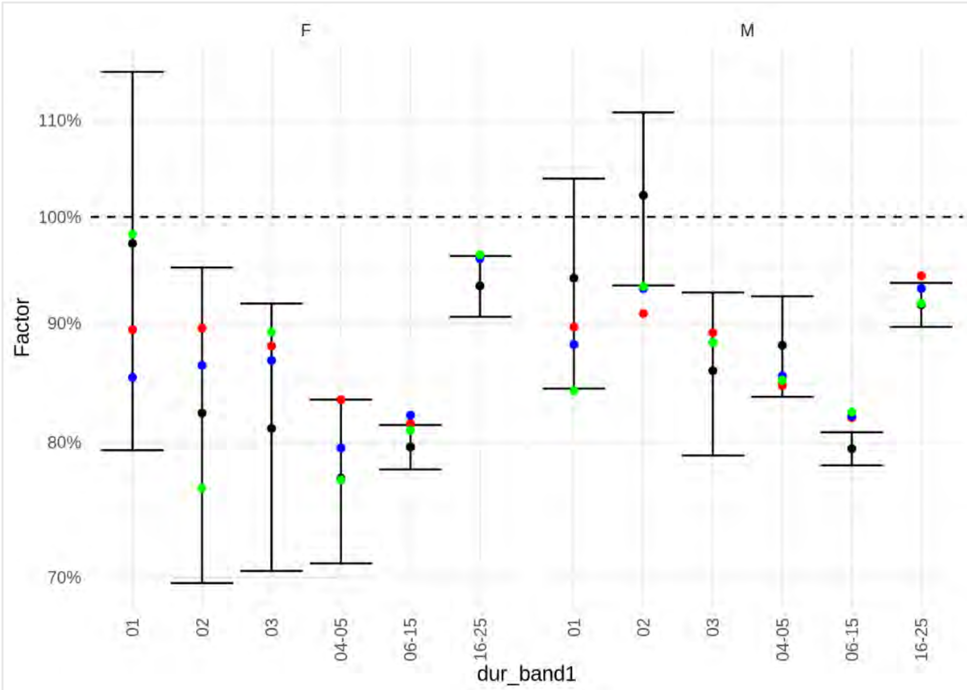
17face_amount_band by insurance_plan



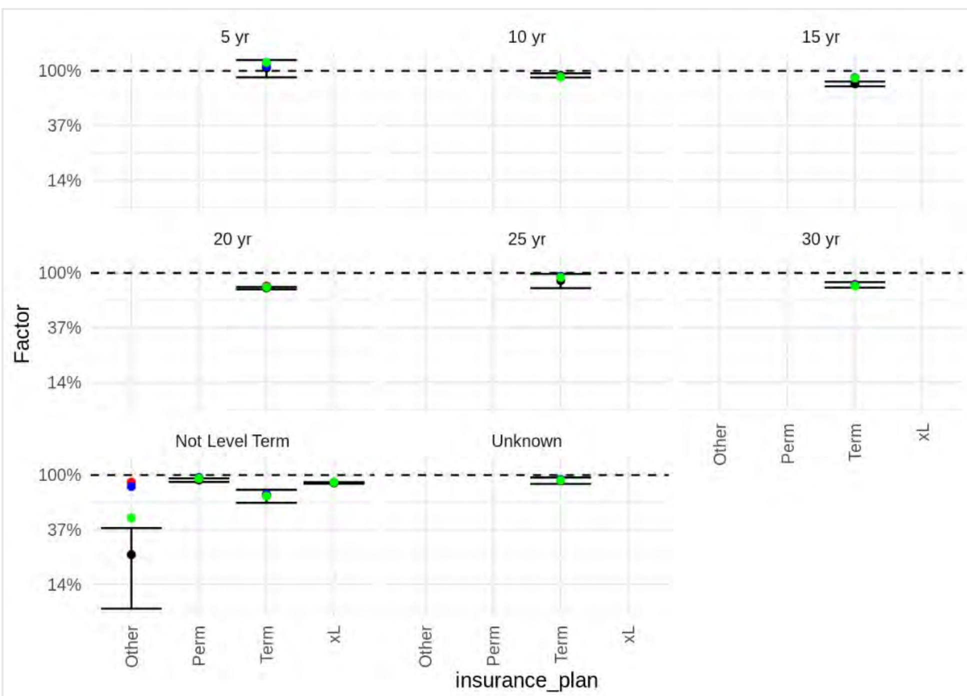
18ia_band1 by insurance_plan



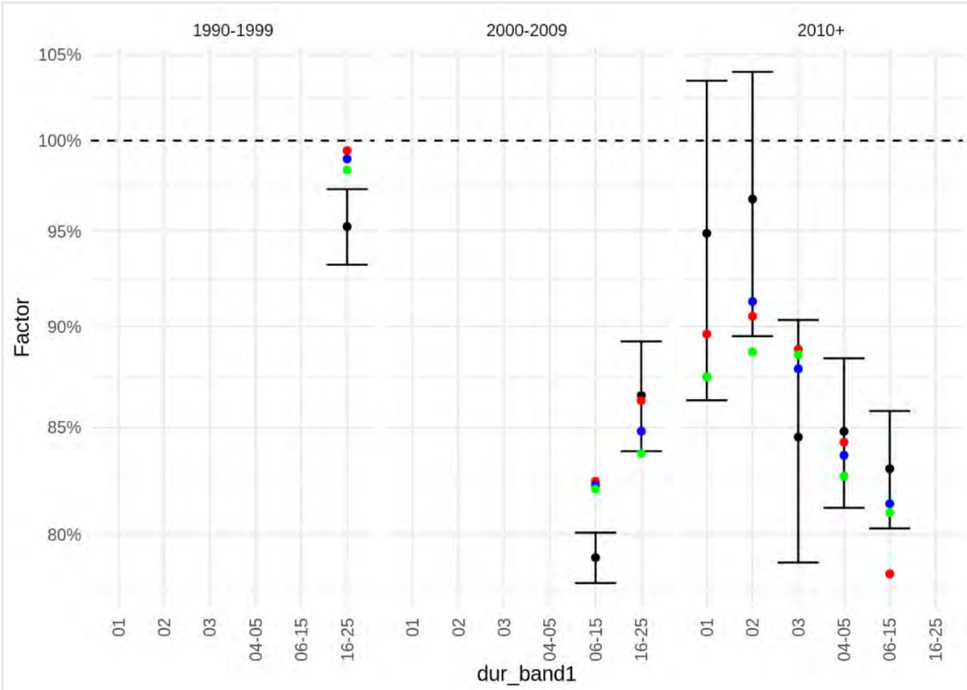
19iy_band1 by uw



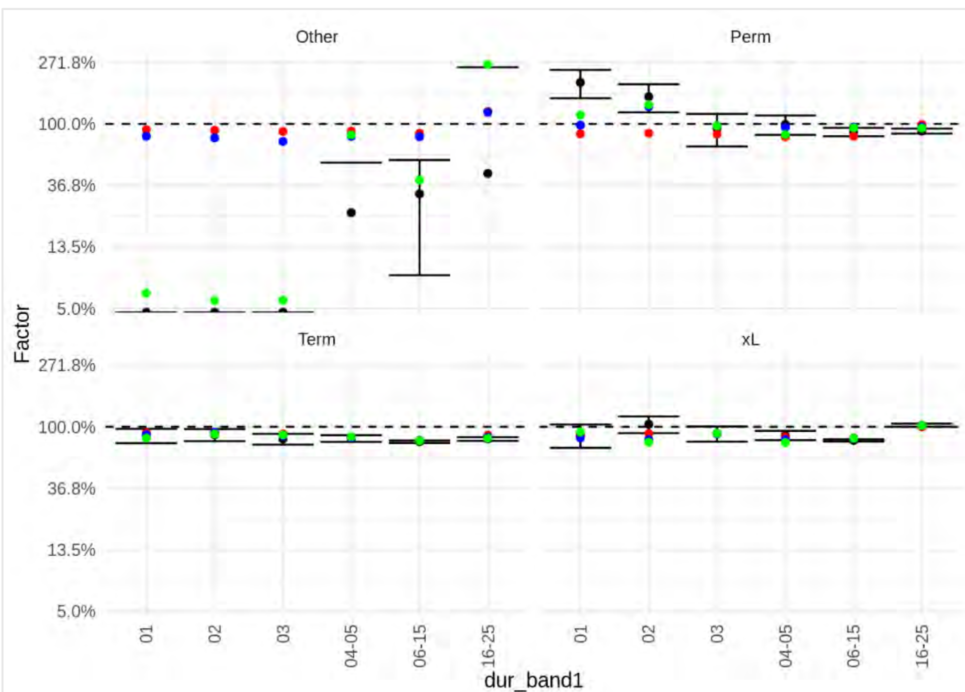
20dur_band1 by gender



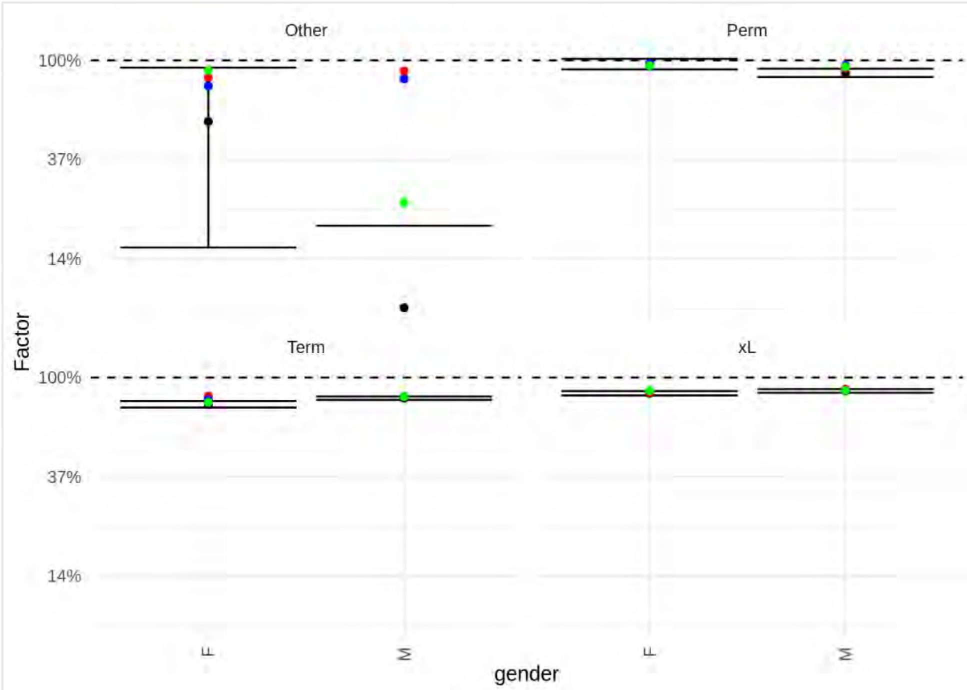
21insurance_plan by ltp



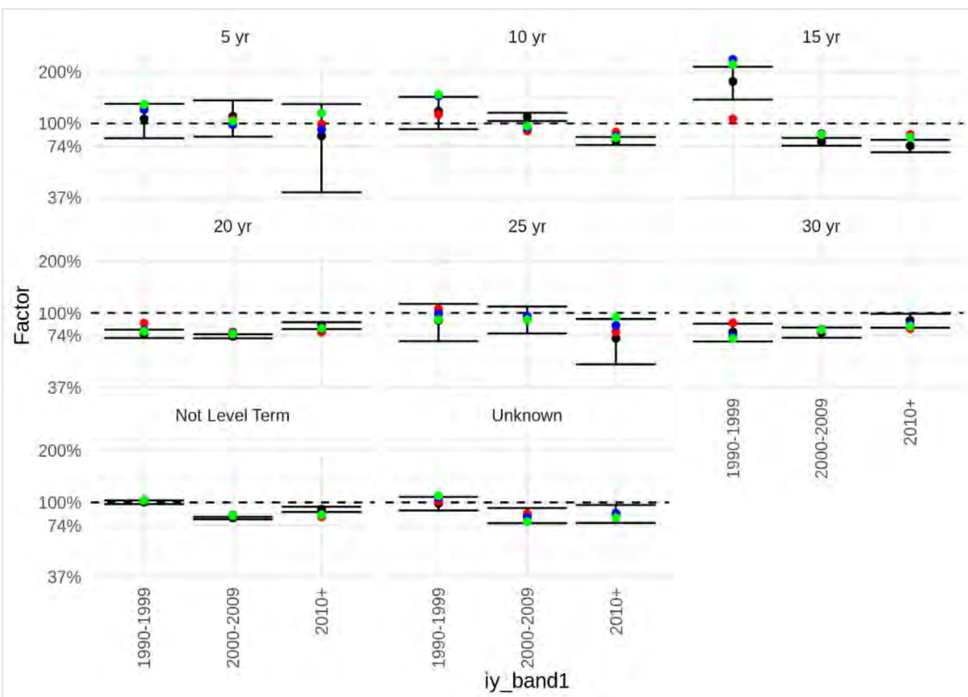
22dur_band1 by iy_band1



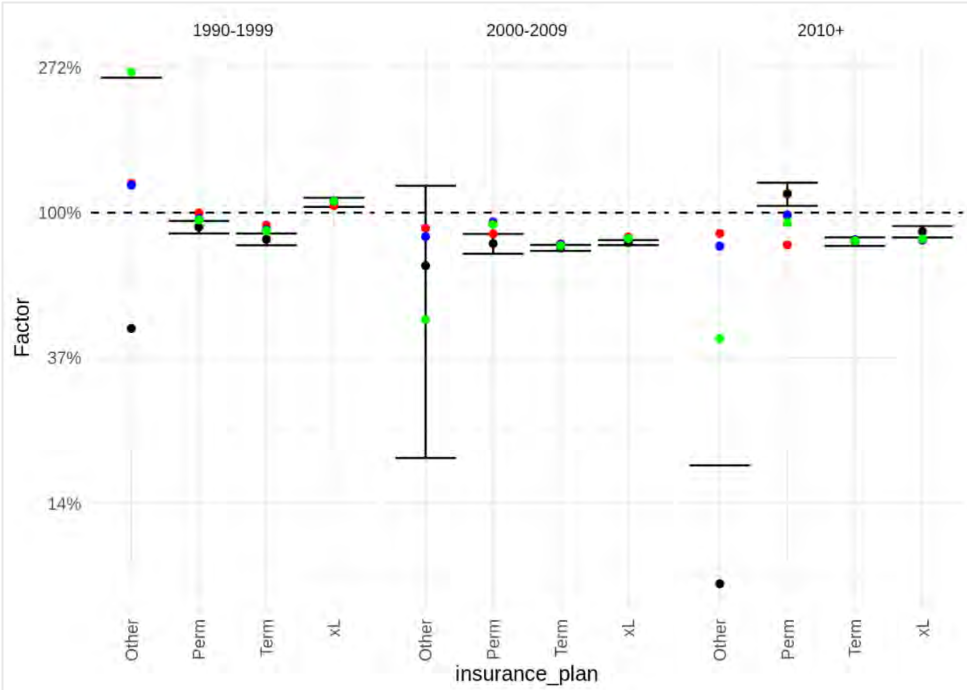
23dur_band1 by insurance_plan



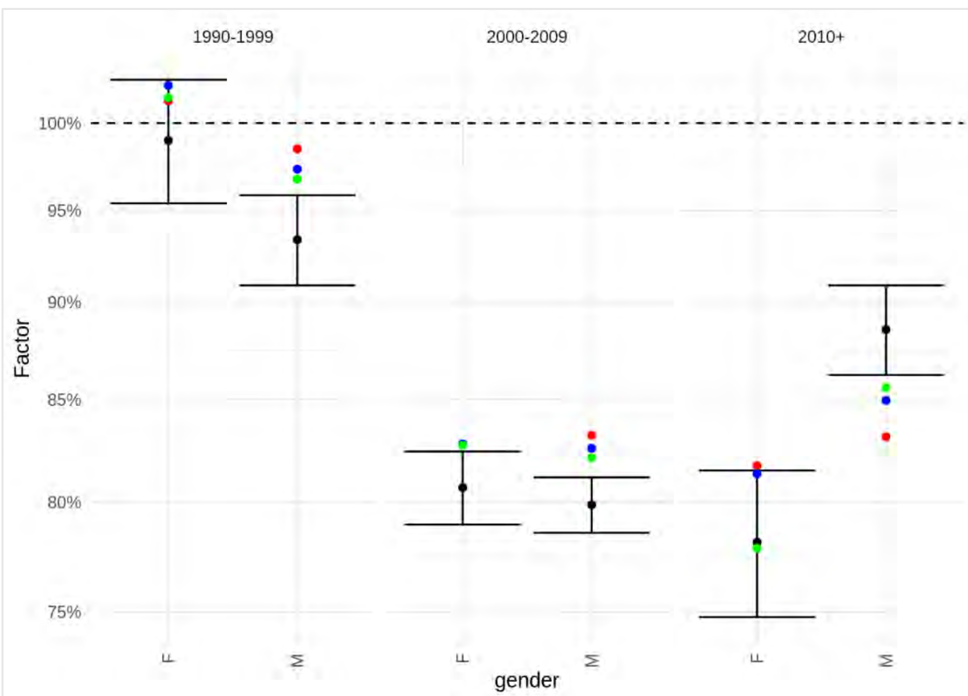
24gender by insurance_plan



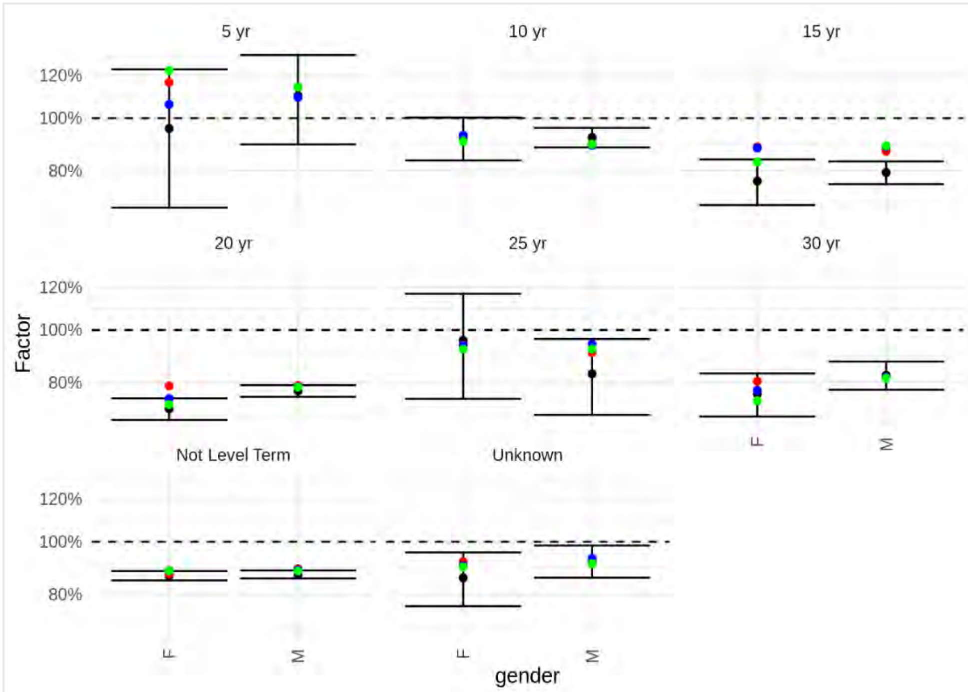
25iy_band1 by ltp



26insurance_plan by iy_band1



27gender by iy_band1



28gender by ltp

References

Corporation, Microsoft, and Steve Weston. 2022. *doParallel: Foreach Parallel Adaptor for the 'Parallel' Package*. https://CRAN.R-project.org/package=doParallel.

Dowle, Matt, and Arun Srinivasan. 2023. *Data.table: Extension of 'Data.frame'*. https://CRAN.R-project.org/package=data.table.

Fokkema, Marjolein. 2020. "Fitting Prediction Rule Ensembles with R Package pre." *Journal of Statistical Software* 92 (12): 1–30. https://doi.org/10.18637/jss.v092.i12.

Gohel, David, and Panagiotis Skintzos. 2023. *Flextable: Functions for Tabular Reporting*. https://CRAN.R-project.org/package=flextable.

Maksymiuk, Szymon, Ewelina Karbowiak, and Przemyslaw Biecek. 2021. *EIX: Explain Interactions in 'XGBoost'*. https://CRAN.R-project.org/package=EIX.

Mayer, Michael. 2023. *Shapviz: SHAP Visualizations*. https://CRAN.R-project.org/package=shapviz.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Pedersen, Thomas Lin. 2023. *Patchwork: The Composer of Plots*. https://CRAN.R-project.org/package=patchwork.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoş Moldovan-Grünfeld, Jeroen Ooms, and Apache Arrow. 2023. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

Shi, Yu, Guolin Ke, Damien Soukhavong, James Lamb, Qi Meng, Thomas Finley, Taifeng Wang, et al. 2023. *Lightgbm: Light Gradient Boosting Machine*. https://CRAN.R-project.org/package=lightgbm.

Tay, J. Kenneth, Balasubramanian Narasimhan, and Trevor Hastie. 2023. "Elastic Net Regularization Paths for All Generalized Linear Models." *Journal of Statistical Software* 106 (1): 1–31. https://doi.org/10.18637/jss.v106.i01.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Maximilian Girlich, Mark Fairbanks, and Ryan Dickerson. 2023. *Dtplyr: Data Table Back-End for 'Dplyr'*. https://CRAN.R-project.org/package=dtplyr.

Zeileis, Achim, and Torsten Hothorn. 2002. "Diagnostic Checking in Regression Relationships." *R News* 2 (3): 7–10. https://CRAN.R-project.org/doc/Rnews/.

About The Society of Actuaries Research Institute

Serving as the research arm of the Society of Actuaries (SOA), the SOA Research Institute provides objective, data-driven research bringing together tried and true practices and future-focused approaches to address societal challenges and your business needs. The Institute provides trusted knowledge, extensive experience and new technologies to help effectively identify, predict and manage risks.

Representing the thousands of actuaries who help conduct critical research, the SOA Research Institute provides clarity and solutions on risks and societal challenges. The Institute connects actuaries, academics, employers, the insurance industry, regulators, research partners, foundations and research institutions, sponsors and non-governmental organizations, building an effective network which provides support, knowledge and expertise regarding the management of risk to benefit the industry and the public.

Managed by experienced actuaries and research experts from a broad range of industries, the SOA Research Institute creates, funds, develops and distributes research to elevate actuaries as leaders in measuring and managing risk. These efforts include studies, essay collections, webcasts, research papers, survey reports, and original research on topics impacting society.

Harnessing its peer-reviewed research, leading-edge technologies, new data tools and innovative practices, the Institute seeks to understand the underlying causes of risk and the possible outcomes. The Institute develops objective research spanning a variety of topics with its strategic research programs: aging and retirement; actuarial innovation and technology; mortality and longevity; diversity, equity and inclusion; health care cost trends; and catastrophe and climate risk. The Institute has a large volume of topical research available, including an expanding collection of international and market-specific research, experience studies, models and timely research.

Society of Actuaries Research Institute
8770 W Bryn Mawr Ave, Suite 1000
Chicago, IL 60631
www.SOA.org