Maryland
Bristol West Insurance Company
SERFF: BRWS-134880995

EXPLANATORY MEMORANDUM

Bristol West Insurance Company is filing to introduce a new Private Passenger Automobile program in Maryland. The proposed effective date for new business is June 24, 2026. The accompanying forms are filed under BRWS-134791685.

The following items are included in this Rate/Rule Filing:

Rate/Rule Schedule

- Rate Manual
- Rule Manual
- UW Guidelines
- Vehicle Symbol Manual
- Credit Model – Filed Confidentially

Supporting Documents

- MD BAC 4.1 – Actuarial Filing Support_Credit
    - Contains a Data Dictionary and Actuarial and Statistical Support for the Credit Model
- Actuarial Memorandum and Support for Rate Factor Selections

Bristol West Insurance Group
Actuarial Memorandum – Modeling Process

*** General Modeling ***

The Rate Order of Calculation was built by line coverage using particular cases of Generalized Linear Models (see, for example, McCullagh and Nelder, 1989). In brief, the Generalized Linear Models approach allows a flexible model form that permits a variety of link functions and exponential error distributions. This flexibility is useful to more accurately adapt models to the data as a means of enhancing pricing precision.

The factors for the Rate Order of Calculation were obtained by modeling pure premium directly, including the bucketed credit scores as a predictor. Because of this, the correlation between all effects is controlled, including (for example) the correlation between Symbol and Model Year/Vehicle Age. The specific form of the ROC is hybrid, with the majority of the effects being multiplicative but Driver Class, Points, and Violation Aging additive. Both the multiplicative and additive portions of the model were estimated using a Generalized Linear Model with a logarithmic link function and Poisson error. The earned car years were used as a weighting variable.

The sheer number of model parameters demands the modeling be conducted in steps due to convergence issues that are typically encountered when fitting a full model. Credit Tier was calibrated last given the MD state specific restrictions. Mathematically, the steps are solved as follows:
Assume that the modeling dataset has J observations and the observed pure premium for the j-th observation is PP_j, j=$\overline{1, J}$. Define the predicted pure premium for each observation as P_PP_j, where again j=$\overline{1, J}$. Assume that there are N predictors and predictor i has n_i levels for which n_i coefficients are estimated. The GLM can be written as:

$$\ln (P\_PP_j) = \beta_0 + \sum_{l=1}^{N} \sum_{k_l=1}^{n_l} \beta_{kl} * \delta_{kl}$$

The log-likelihood that is minimized is the Poisson log-likelihood:

$$LL\ (\boldsymbol{\beta}; \mathbf{PP}) = -\sum_{j=1}^{J} ECY_j * (PP_j * \ln (P\_PP_j) - P\_PP_j)$$

where:

$\boldsymbol{\beta}$ is the vector of the estimated coefficients, $\mathbf{PP} = (PP_1, PP_2, \ldots , PP_J)$ is the vector of the actual pure premium and δ_{kl}, are indicators that declare if observation j belongs to the level k_l of the i-th variable.
The coefficients are the result of the minimization process of the likelihood. That is solved iteratively inside the HPGENSELECT procedure in SAS through an algorithm like Newton – Raphson, Conjugate Gradient, etc. Please see the SAS documentation for procedures like GENMOD and HPGENSELECT here:
https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.3/statug/statug_genmod_details01.htm
for the theory, and here:
https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.4/statug/statug_hpgenselect_syntax01.htm#statug.hpgenselect.proc_technique
for the various techniques used by SAS to estimate the coefficients.

Bristol West Insurance Group
For the direct application of GLM models in insurance, please consult "GENERALIZED LINEAR MODELS FOR INSURANCE RATING" found here:
https://www.casact.org/sites/default/files/2021-01/05-Goldburd-Khare-Tevet.pdf

After estimation, we reach the final form of the factors as follows:

a. Exponentiation – the model is multiplicative with log link
b. Rebasing – the exponentiated factors are rebased to an exposure weighted average of 1.00 or to a given level depending on case

The mechanics of model solving in steps are straightforward. Some effects are modeled, a residual is computed, and additional effects are then modeled on the residual. In particular, we sought to reduce the number of correlated effects solved in a particular step to mitigate rate dispersion upon a subsequent factor refresh. As an example, there is a fair amount of correlation between Vehicle Type and Symbol. From a precision standpoint, it doesn't matter which effect is modeled first, but modeling both effects simultaneously could produce noticeable factor changes when new data is added. This occurs as one effect "flattens" the other effect, then vice versa for a new dataset. In our experience, overall precision is unchanged, but rates are altered.

The ROC modeling procedure relies on exposures (earned car years) as weights, incurred losses and incurred claim counts of our existing policyholders to develop indicated relativities. The modeling dataset consists of loss experience from all states, excluding CA for the period 06/01/2018 to 05/31/2023 evaluated as of 12/01/2023. Losses for BI, PD, CM, CL and PIP were developed to ultimate. The data included all products except for Basic and Prima. For modeling, we used a random 70% sample of the data and tested the models on the remaining 30% hold-out data. Catastrophe losses were eliminated.

*** LS MEANS ***

In order to avoid excessive rate, as a final step of the modeling process, we complete the LSMEANS process. For LSMEANS, R&D scores the countrywide data with the selected factors and then inspects the ratio of the observed pure premium divided by the fitted pure premium for every level of every rating variable. This "Ratio Residual" is thus defined as:

Ratio Residual = Observed_PP / Predicted_PP.

These ratio residuals are then plotted for every level of every rating variable used in the ROC. Within any given rating variable, if the product is perfectly priced, the ratio residuals should have a slope of zero and a weighted average of exactly 1.0. To the extent that the process of selecting factors distorted the model, the ratio of observed pure premium to fitted pure premium may display nonzero slope within a rating variable. As long as one variable at a time is modified, the ratio of observed to fitted values can be used as a correction factor to "fine tune" the selects and improve accuracy. Specifically, the formula is:

New ROC Factor = Ratio Residual * Original Selected Factor.

The selection of the variables to be modified in the LS MEANS is based on their potential to improve accuracy. A variable that has a set of ratio residuals that are overall further from 1.00 is considered to be a stronger candidate for model improvement. When a variable is modified we are careful to avoid generating illogical rates. For example, if we alter a variable like Credit Tier we take into account the fact that this variable needs to be monotonically increasing.

*** Driver Class Extrapolation ***

We group driver class ages to aid both model convergence and credibility when selecting factors. Unfortunately, this means that we do not have a parameter estimate for every specific year of age. To obtain smooth yearly price transitions as a driver ages, we need to interpolate/extrapolate the modeled age groups across the age span from 14 to 99. We conduct these analyses separately by coverage, and within coverage, separately for each classification of marital status/gender.

We compute the weighted average age for each of our grouped age categories. We defined these weighted averages as fixed knot values in spline regression analyses. Spline regression provides a smoothing function through the modeled age groups that is more flexible than a polynomial regression approach. The dependent measure in the spline regression is the original model selected relativity for the age group weighted by earned exposures.

The output from the spline regression provides predicted relativities for each age between 14 and 99. Unfortunately, these predictions only roughly conform to a logical requirement that the weighted average prediction for an age group equals the original model selected relativity for that age group. To enforce this requirement, the spline regression predictions were post-processed using a constrained estimation algorithm. The target of the algorithm was the predicted value from the spline regression (weighted by allocated earned exposures). The final coefficients are calculated such that their weighted average for an age group exactly equals the pure premium model selected factors for that age group.

*** Points Extrapolation ***

Much like Driver Class, we group Points into different clusters for modeling purposes. Grouping will help us achieve better credibility in our indicated factors since there will be more claims in the grouped Points versus the ungrouped Points. The grouping of Points is different for each coverage. To interpolate Points, we used the factors from modeling and ran monotonic splines through these factors using the exposures of the raw Points variable. Beyond the left endpoint of the last Points cluster, we used linear extrapolation.

*** Vehicle Age ***

In the modeling process Vehicle Age is bounded at 22. Vehicle Ages lower than 22 are used as individual levels. For production we expand the Vehicle Age to 30 by using linear extrapolation.