



COLORADO

Department of
Regulatory Agencies

Division of Insurance

Marguerite Salazar
Commissioner of Insurance

Request for Generalized Linear Model Details

The following technical information must be provided in any filing using generalized linear models (GLMs) in the development of proposed rates and rating factors.

Raw Input Data

- Provide the data fields (premiums, losses, loss ratios etc.) that were used as inputs in the models.
- List all policyholder variables/characteristics used in each model, and provide a complete description for each to include:
 - (1) A definition for the variable, indicating all levels of the variable used – what type of variable it is: continuous, ordinal, nominal
 - (2) A detailed description of how each variable is used in each of the GLM models;
 - (3) Source of data used for the variable, whether internal company data, or external data
 - (4) A list of rating variables that were originally considered but ultimately rejected and the reasons why they were rejected.
 - (5) Are the variable used loss-related, expense-related or related to the risk in some other way?
- Indicate if the input data is based solely on Colorado experience. If so, was the model validated using Colorado data? If not, please include any other internal or external data sources used in the models. Discuss the credibility of these data sources.
- Indicate exposure periods included in the data, how often the data is updated in the models.
- Discuss whether any adjustments were made to the input data.
- Disclose whether catastrophic data was included or excluded in the underlying data used to develop the models. If included, then explain how this was appropriate and what technique was applied to minimize its effects on the models results.

Structure of GLMs

- Indicate the software used in constructing the models. If an outside proprietary model is used indicate the degree to which the company may modify model parameters.
- Is the company using multiple GLMs? If so, provide the model specs for all GLMs and demonstrate how they work together to determine the pure premium for each



policyholder. List all the assumptions and methodologies used in the structure of all GLMs.

- Include all mathematical formulas used in the models.
- Describe any judgment employed throughout the modeling process.
- Indicate whether the company chooses which policyholder characteristics can be turned off/on to rate a specific policyholder. Include all criteria for when and why this is done.
- Indicate the target function for each GLM, any maximization or minimization process used, with constraint conditions.
- Clearly indicate all capping rules and formulas, if they are set subjectively, indicate the plan to move to indicated rates over time.
- Indicate in detail the differences between rating new policyholders versus renewing policyholders.
- Indicate how all rate range confidence intervals are developed in the models for actuarially sound rate ranges between the indicated and selected rates. Provide all parameters used in setting confidence intervals, such as 90% confidence level, etc...
- Specify the levels of characteristics and interactions between the variables in the models.
- Provide the actual LINK function, the inverse of which transforms the linear combinations of observed values of independent variables (explanatory factors or predictor variables) into expected values of the response variable (frequency, severity, or pure premium). Provide the distribution used for the error terms (frequency, severity, or pure premium).
- Identify any constrained variables and their minimum and maximum values.
- Identify any fixed variables and the reasons why they are fixed within the models.
- Indicate test performed to verify that the models satisfy any of the following that are applicable, the degree to which they are applicable - indicate any of these that are not applicable or are violated:
 - (1) Tests for independence of observations, independence of residuals;
 - (2) Linear relationship between dependent variable and each independent variable;
 - (3) Test for Homoscedasticity;
 - (4) Test for Multi-collinearity;
 - (5) Test for significant outliers;
 - (6) Check that residuals are approximately normally distributed;
 - (7) Any other testing that is done on your models.
- Discuss tests for and adjustments made for correlation.

Model Outputs

- Provide all statistical measures to indicate the model predictive value and goodness-of-fit (R, R-squared, Adjusted R-squared, Standard Error, Model F-Ratio and its significance).

- Provide the output coefficient table showing the coefficients for each variable used in the models. This should typically include for each variable:
 - (1) Unstandardized coefficient indicating how much the dependent variable varies with the independent variable when all other independent variables are held constant;
 - (2) Standardized coefficient beta;
 - (3) Confidence interval for each coefficient;
- Provide statistical summaries to indicate that each variable in the models is statistically significant and is predictive for expected losses or expenses (F-test, t-test, p-values, other). If the variable is not predictive of loss or expense indicate in what way it is related to risk.
- Indicate the process and all measures used for introducing and removing variables in modeling.
- What is the purpose of the model used? What does the model seek to maximize or minimize (e.g., underwriting profit, retention, other) and explain. Discuss whether the goal of the model is price optimization.
- Provide a comparison, in an excel file, of the present, model indications and proposed rating factors. Include support for instances where the indicated and proposed rating factors differ.

The Division would allow the detailed rating information described above to be provided as Confidential. If the company requests confidentiality for any filing documents, a properly completed Confidentiality Index must be submitted. Refer to Bulletin B-1.15 regarding the Confidentiality Index and confidential information. Confidential documents must be separated from non-confidential documents and they need to be marked with the confidential icon as well. The Confidentiality Index itself is not confidential.