APPENDIX B-RGLM – INFORMATION ELEMENTS AND GUIDANCE FOR A REGULATOR TO MEET BEST PRACTICES' OBJECTIVES (WHEN REVIEWING REGULARIZED GENERALIZED LINEAR MODELS)

and does not necessarily need to be included by the filer with the initial submission, unless specifically requested by a particular state. It is typically requested only if the reviewer has serious concerns that the model may produce rates or rating factors that are excessive, inadequate, and/or unfairly discriminatory.

Appendix B-RGLM is focused on Regularized GLMs including lasso, derivative lasso, lasso credibility, ridge, elastic net, and accurate generalized linear models. This appendix should not be referenced in the review of other model types. This Appendix RGLM is intended to provide state guidance for the review of rate filings based on regularized GLMs.

A. SELECTING MODEL INPUT

Section	Information Element	Level of Importance to the Regulator's Review	Comments
1. Available Data Sources			
A.1.a	Review the details of sources for both insurance and non-insurance data used as input to the model (only need sources for filed input characteristics included in the filed model).	1	Request details of data sources whether internal to the company or from external sources. For insurance experience (policy or claim), determine whether data are aggregated by calendar, accident, fiscal, or policy year and when it was last evaluated. For each data source, get a list of all data elements used as input to the model that came from that source. For insurance data, get a list all companies whose data is included in the datasets.
			Request details of any non-insurance data used (customer-provided or other), whether the data was collected by use of a questionnaire/checklist, whether data was voluntarily reported by the applicant, and whether any of the data is subject to the federal Fair Credit Reporting Act (FCRA). If the data is from an outside source find out what steps were taken to verify the data was accurate, complete, and unbiased in terms of relevant and representative time frame, representative of potential exposures, and lacking in obvious correlation to protected classes.
			Note: Reviewing source details should not make a difference when the model is new or refreshed; refreshed models would report the prior version list with the incremental changes due to the refresh.
			Accuracy of insurance data should be reviewed. It is

A.1.b	Reconcile aggregated insurance data underlying the model with available external insurance reports.	4	

Section	Information Element	Level of Importance to the Regulator's Review	Comments
A.1.c	Review the geographic scope and geograp exposure distribution of the raw data for relevance to the state where the model is filed.	2	

Section	Information Element	Level of Importance to the Regulator's Review	Comments
			To accelerate the review of the filing, it may be desirable to request (from the company) the name and contact information for a vendor representative. The company should provide the name of the third-party vendor and a contact in the event the regulator has questions. The "contact" can be an intermediary at the insurer (e.g. <(pa)4s(nf)1c 0 Tw 2.952 0 Td2.952 0 T
A.2.c	Determine if the sub-model output was used as input to the Regularized GLM; obtain the vendor name, as well as the name and version of the sub-model.	1	

A.3.c Ask for aggregated data (one dataset of- pre adjusted/scrubbed data and one dataset of- post adjusted/scrubbed data) that allows the regulator to focus on the univariate distributions and compareenaocus.2 (t)64.2 (omw1.6 (o5 (n)-7 (a6.9 (of)1.7w1.6 (o5 (us)9.4

A.4.c	Identify material findings the company had during its data review and obtain an explanation of any potential material limitations, defects, bias, or unresolved concerns found or believed to exist in the data. If issues or limitations in the data influenced modeling analysis and/or results, obtain a description of those concerns and an explanation of how modeling analysis was adjusted and/or results were impacted.	1	"None" or "N/A" may be an appropriate response.

Section	Information Element	Level of Importance to the Regulator' s

| B.1.h | Obtain a description of the candidate variable selection process prior to the model building. | 1 |

Candidate variables are the variables used as input to the modeling process. Certain variables may not end up used in the final model as some regularized GLM models (lasso, elastic net, etc.) will remove less significant variables. The narrative regarding the candidate variable selection process may address matters such as the criteria upon which variables were selected or omitted, identification of the number of preliminary variables considered in developing the model versus the number of variables that remained, and any statutory or regulatory limitations that were taken into account when making the decisions regarding candidate variable selection.

The modeler should comment on the use of automated feature selection algorid ( )Tj.892 -1.157 Td [(r)-2.3 (eg)-4.

3. Predictor Variables

Obtain a complete data dictionary, including the
names, data types, definitions, and uses of ea...
predictor variable, offset variable, ...

B.3.a

4. Adjusting Data, Model Validation, and Goodness-of-Fit Measures			
B.4.a	Obtain a description of the methods used to assess the statistical significance/goodness-of-fit of the model to validation data, such as lift charts a statistical tests. Compare the model's projected results to historical actual results and verify that modeled results are reasonably similar to act results from validation data.	1	For models that are built using multistate data, validation data for some segments of risk is likely to have low credibility in individual states. Nevertheless, some regulators require model validation on state only data, especially when analysis using state only data contradicts the countrywide results. State only data might be more applicable but could also be impacted by low credibility for some segments of risk. Note: It may be useful to consider geographic stability measures for territories within the state.

Statistical confidence intervals and p-values are often not available for Regularized GLMs. However, there are other ways to demonstrate model stability. The model could be run 100+ times on bootstrapped datasets to det (e)-2.8s on bootstrapped7l-0 0 9.9 7 (bou)1.

B.4.b For all variables, review the appropriate parameter values and relevant demonstrations of stability. Relevant demonstrations of stability may be provided as either plots by variable of indicated factors which also show upper bound and lower bound values (95th percentile and 5th percentile) on bootstrapped datasets, coefficient ranges across dataset folds, or p-values from a comparable standard GLM.

1

B.4.d	Obtain a description how the model was tested for stability over time.	2	Evaluate the build/test/validation datasets for potential time-sensitive model distortions (e.g., a winter storm in year 3 of 5 can distort the model in both the testing and validation datasets).

Obsolescence over time is a model risk (e.g., old data for a variable or a variable itself may no longer be relevant). If a model being introduced now is based on losses from years ago, the reviewer should be interested in knowing whether that model would be predictive in the proposed context. Validation using recent data from the proposed context might be requested. Obsolescence is a risk even for a new model based on recent and relevant loss data.

The reviewer may want to inquire as to the following: What steps, if any, were taken during modeling to prevent or delay obsolescence? What controls exist to measure the rate of obsolescence? What is the plan and timeline for updating and ultimately replacing the model?

The reviewer should also consider that as newer technologies enter the market (e.g., personal automobile) their impact may change claim activity over time (e.g., lower frequency of loss). So, it is not necessarily a bad thing that the results are not stable over time. |
| B.4.e | Obtain a narrative on how potential concerns with overfitting were addressed. | 2 | |
| B.4.f | Obtain support demonstrating that the overall Regularized GLM assumptions are appropriate. | 3 | A visual review of plots of actual errors is usually sufficient.

The reviewer should look for a conceptual narrative covering these topics: How does this particular Regularized GLM work? Why did the rate filer do what it did? Why employ this design instead of alternatives? Why choose this particular distribution function and this particular link function? A company response may be at a fairly high level and reference industry practices.

If the reviewer determines that the model makes no assumptions that are considered to be unreasonable, the importance of this item may be reduced. |
| B.4.g | Obtain 5-10 sample records with corresponding output from the model for those records. | 4 | |

Section	Information Element	Level of Importance to the Regulator's Review	Comments

Section	Information Element	Level of Importance to the Regulator's Review	Comments

8. Accurate Translation of Model into a Rating Plan

| | Obtain sufficient information to understand how the model outputs are | | |
| C.8.a | | | |