



Cognalysis MultiRate™ Version 4 – Release Notes

August 2017

We are pleased to release Version 4 of Cognalysis MultiRate™.

Major Changes

- Guided characteristic selection
- Layered smoothing
- Bootstrapping of test data
- Improvement modes
- Improved visualization of results
- Shrink zeroes mode

Minor Changes

- Expanded geographical results
- Rebuild .CGMR file using output prefix
- Addition of lakes and rivers to maps
- Interaction effect viewing style
- Open exported files from MultiRate
- Miscellaneous bug fixes

Guided Characteristic Selection

When the “Use” button is checked in the Column Selection screen, MultiRate will attempt to detect the characteristic’s data type.

- Numeric or date fields that take a large number of values will be set to “Grouped”.
- Fields meeting the profile of the specific geographic options will be identified as such.
- Potential detail keys can often be automatically detected.

When the “Select All” button is clicked, MultiRate will automatically perform the characteristic selection process across all of the variables. In addition, fields that take only one value will not be selected.

Layered Smoothing

Layering is being provided as an additional option for smoothing a characteristic. Similar to Variable Gradient Smoothing, multiple radiuses are tested, but selecting “Layered” treats each radius as a separate layer that together are combined to obtain the factor for a given bin/geography. Layered characteristics can be viewed in their individual parts in the result inspection panel, or as an aggregate. Individual layers can be turned off (removed) by right-clicking in the Status of Analysis panel and choosing “Trim Characteristics”, giving the user more direct control over the smoothing. Because each layer is considered distinct, the parameter count is considered higher than a corresponding Variable Gradient selection in which a single radius is selected for each bin/geographical unit. For this reason, the Layered smoothing type will generally result in a higher parameter (Fuzzy K) count.

Bootstrapping of Test Data

When comparing alternative models against each other, a Komogorov-Smirnov test is used to determine whether the results on test data are statistically different. In previous versions, the K-S test compared the error distributions of the two models being compared. It was possible that models could have significantly different error distributions, but that the primary fit statistics of interest were not appreciably different. This can lead to situations where some simpler models are unduly rejected, and even to cases where better fits to the data are not found because the improvement process is curtailed early.

Version 4 improves on this concept by performing the K-S test on the very ratio being used to identify the model with the best fit: Relative Net Lift/Average Absolute Error. To make this comparison, 20 different bootstrapped samples of the test data are made, and this ratio is recalculated for each bootstrapped sample. The K-S test is then performed between the distributions of this ratio for the alternative models.

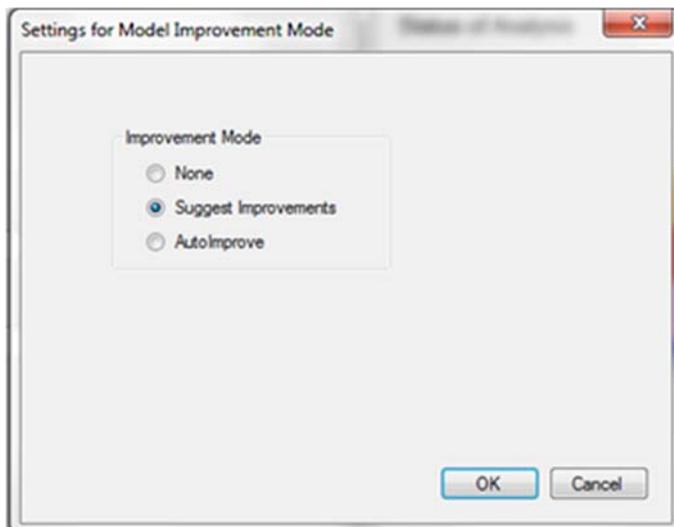
Improvement Modes

An important part of the modeling process is to refine your model to achieve the best predictive power possible, while simultaneously seeking to simplify the model specification. In addition to the ability in previous versions to make adjustments and see the impact of those adjustments, Version 4 can be used to more directly guide the model refinement process itself.

If the Improvement Mode feature is used, MultiRate will consider the following changes:

- **Dropping characteristics** (based on a low ratio of Significance to the square root of Fuzzy K, or on the amount of duplication with other variables)
- **Adjusting credibility** (based upon the # of changes in sign of [actual – modeled] along the modeled curve)
- **Creating characteristics based on interaction effects** (based on a scan of the interaction statistic for all 2-way combinations)
- **Altering characteristic smoothing to Linear** (based on linearity of adjusted factors)
- **Changing bin breakpoints for grouped characteristics** (based on the differences in slope along the curve)

There are two new improvement modes a user may utilize: Suggest Improvements and AutoImprove. As shown in the dialog box below, users will still be able to run an analysis without an improvement mode. Running with “None” selected will not generate any suggestions (such as in previous versions). To select an improvement mode, click the light bulb on the toolbar, select the desired improvement mode and click “OK”.



The light bulb icon will change color to indicate the current improvement mode.

 = None

 = Suggest Improvements

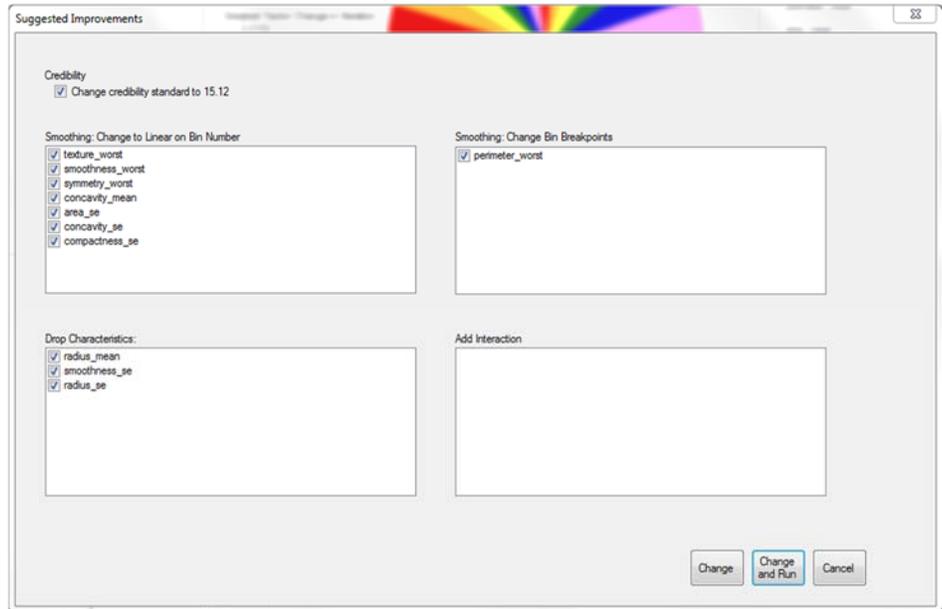
 = AutoImprove

The results on the training data only are used to form the basis for MultiRate’s recommendations.

Suggest Improvements Mode

In this improvement mode, MultiRate will suggest adjustments, and then display a dialog box containing those potential adjustments. This will occur immediately after running an analysis and also when the user presses the start button from an existing analysis that has been completed.

The user can then select which improvements to implement.



After the desired adjustments have been selected, there are three options to proceed: Change, Change and Run, or Cancel. Clicking the “Change” button will implement the selected adjustments and return to the main screen, letting the user implement any additional changes. “Change and Run” will implement the changes and immediately start a new analysis. The new analysis will not overwrite the previous one. Lastly, “Cancel” will implement no changes, regardless of what is chosen, and return to the main screen.

Improvements during the same MultiRate session to subsequent alternatives for a model will retain a memory of previous adjustments, avoiding the potential for loops in the improvement process. In some cases the suggested improvement will be to return to a previous model and start from that point with more moderate changes.

After each revised model has finished running, the status bar at the lower left of the screen will indicate which model has the best fit on the test data, and which model is the simplest within tolerance of the best fit.

In the final stages of the improvement process, MultiRate suggests ever more stringent credibility standards as a final method of increasing model robustness.

MultiRate eventually reaches a state where none of the potential suggestions are considered likely to improve the model by giving it a better fit or by making it simpler and therefore no further suggestions are offered.

AutoImprove Mode

AutoImprove Mode follows the same process as Suggest Improvement Mode, as if “Change and Run” were selected after each analysis. The process continues until no more improvements are suggested. The alternative versions are each saved (and renumbered) without the need for approval by the user.

Results can still be inspected while the process continues, and the status bar will continue to update with information about the best and simplest analyses.

Optional functionality when running AutoImprove includes the choice of whether to add interaction effects. Interaction effects can give an advantage in improving model fit, but they do add a degree of model complexity.

Another option that is presented to the user with AutoImprove Mode is whether to delete extraneous analyses. This can be useful for file/space/database management. When running on AutoImprove, MultiRate may create many analyses. Choosing “Delete extraneous analyses at the end” will tell MultiRate to delete all analyses in the series that are not the best fit, simplest, or first analysis.

Halt Button

The user should be aware that clicking the “Halt” button while an analysis is running results in different behavior depending upon whether an Improvement Mode has been selected or not. If the Improvement Mode is set to “None”, clicking Halt will cause MultiRate to finish the analysis with the current iteration, and not proceed further. If the Improvement Mode is set to “Suggest Improvements” or “AutoImprove”, MultiRate will finish the analysis with the current iteration and then proceed to offer suggestions for the next analysis or begin running the next analysis.

If an Improvement Mode has been selected and the user wishes to halt the entire process (for example, to make corrections to the underlying data), the user should change the Improvement Mode to “None” while the analysis is running and then click the Halt button.

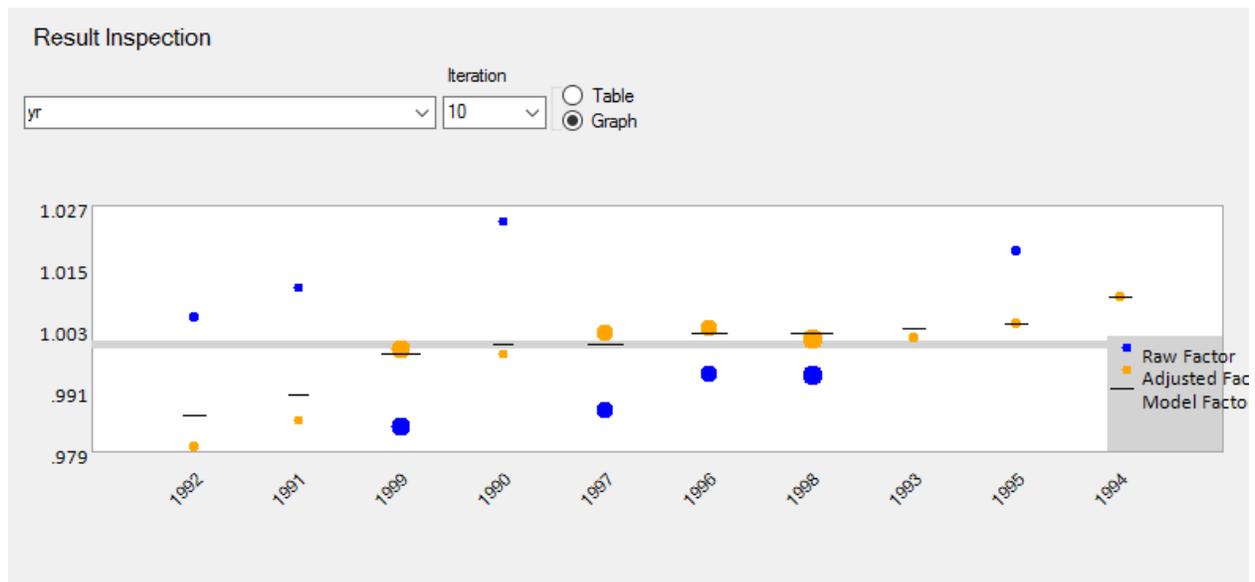
Improvement Log

The improvement log is generated when a series of analyses are run in either Suggest Improvements or AutoImprove mode. The improvement log documents changes as they are made, giving the user a comprehensive view of the changes made after every analysis and the resulting fit statistics and statistical similarity. The user can access the log by selecting File -> Print Improvement Log, then selecting the analysis prefix name from the drop down. Some items in the drop down may say “archived”, those are logs associated with an analysis improvement series that has run more than once.

Result Inspection Panel Improvements

The following changes have been made to the graphical results:

- The dot size is now proportional to the amount of exposure
- The modeled factor is no longer represented by a continuous line unless the bins are based on grouped data or there is smoothing being used
- The factor for the “null” bin will only be shown if the bin contains data
- For categorical, unsmoothed data, results are now shown from low factor on the left to high factor on the right.



“Shrink Zeroes” Mode

A new feature to speed up an analysis is the “Shrink Zeroes” function.

In the case where for the target variable:

- There are more than 3 times as many zero records as non-zero records
- The average value of the target variable is less than 0.5

MultiRate will provide the option to “Shrink Zeroes” in the Iteration Control box. The purpose of this feature is to reduce the time necessary for an analysis in cases with many zero value records, such as a binary model with low probability of a 1 value or an insurance frequency model for smaller accounts with low probability of a claim.

In this approach, a portion of the “zero” records is randomly ignored from the training data. The analysis then proceeds as normal. Rescaling to reincorporate the impact of the excluded records is accomplished through the normal curve modification process. The implicit assumption is that a lower value expectation value for the target variable corresponds to a higher probability of a zero being observed for that variable. If this is the case, the ideal ordering of the characteristic combinations in the original data is identical to the ideal ordering of the characteristic combinations in the reduced data (positive monotonic transformation). Curve modification then estimates the unmodified value given the ordering based on looking at the appropriate cohort of records along the curve for the training data.

Minor Updates

Expanded Geographical Results

In prior versions, only those states, counties or ZIP Codes that were included in the training data received modeled factors. Now factors are provided for all areas within the area bounded by the training data. This is simply an extension of the Variable Gradient/Layered concept.

Rebuild .CGMR file using Output Prefix

If a settings file (.cgmr file) has been deleted, this tool can rebuild it from information stored in the SQL database. By selecting Tools -> Rebuild .cgmr file using Output Prefix, then entering the name of the deleted analysis prefix, MultiRate will attempt to rebuild the settings file. If the rebuild is successful, the user will select a new location to save the .CGMR file.

Map Improvements

Maps of the U.S. now include major rivers and lakes. This additional frame of reference is particularly useful when viewing small areas where other landmarks such as state boundaries or coastlines are not in the viewable area.

Interaction Effects Improvements

The default for viewing interaction effects will now be a list instead of a grid. The list will order interactions from highest interaction score to lowest. The user can view the grid by selecting “Cross View” instead of “List View”.

Open Exported Files from MultiRate

When factors, training data, or test data are exported, the user is now given the option to open the file in MS Excel after it is created.