VR-1: GLMs for Incomplete Development Triangles

Thomas Hartl, Bryant University

Antitrust Notice

- The Casualty Actuarial Society is committed to adhering strictly to the letter and spirit of the antitrust laws. Seminars conducted under the auspices of the CAS are designed solely to provide a forum for the expression of various points of view on topics described in the programs or agendas for such meetings.
- Under no circumstances shall CAS seminars be used as a means for competing companies or firms to reach any understanding – expressed or implied – that restricts competition or in any way impairs the ability of members to exercise independent business judgment regarding matters affecting competition.
- It is the responsibility of all seminar participants to be aware of antitrust regulations, to prevent any written or verbal discussions that appear to violate these laws, and to adhere in every respect to the CAS antitrust compliance policy.

Overview

- There are two sessions:
  - Part I, Model Design (8:30am)
    - Understanding basic set-up and four different model structures for triangle GLMs; includes models with payment period parameters.
  - Part II, Reserve Projections and Ranges (10:30am)
    - Understanding how GLMs can be used to project reserves and measure the variability of reserve outcomes; includes discussion of extending payment period parameters.

Overall Goals

- Understand basic set-up for triangle GLMs
- Explore four different GLM structures for incomplete development triangle
- Understand nature of GLM reserve projections
- Understand nature of bootstrapping ranges
- Understand that reserve projections based on GLMs may be biased due to extrapolation of payment period parameters

Session Format

- Both sessions are interactive and the participants will explore the topics hands-on with the aid of two accompanying MS Excel files:
  - CLRS 2013 - VR1 - Basic Model Structure.xlsx
  - CLRS 2013 - VR1 - Reserves and Ranges.xlsm
- You will need MS Excel 2007 or higher to open the files. Additionally, VBA macros need to be enabled to make use of the second file.

Part I
Model Structure

- Learning Objectives
  - Setting up a model with parameters for each development and exposure period
  - Adding payment period parameters
  - Reducing number of parameters using trend groups
  - Dealing with missing data points
  - Knowing how many data points are needed for a reserve projection model

Setting up a model with parameters for each development and exposure period

- Multiplicative model with parameter for each exposure and development period
- “Slack factor” means that we can drop one parameter without loss of generality
- Linearize model by applying log transform
- Log-linear model can be represented by parameter vector and model matrix

Setting up a model with parameters for each development and exposure period

- Explore these concepts by studying the first three tabs of “CLRS 2013 - VR1 - Basic Model Structure.xlsx”
- Note that at this point we are not fitting a model to data. We are simply studying the structure of the model.

Setting up a model with parameters for each development and exposure period

- Model is linear on log scale:

γ	$\gamma+\beta_2$	$\gamma+\beta_3$	$\gamma+\beta_4$	$\gamma+\beta_5$
$\gamma+\alpha_2$	$\gamma+\alpha_2+\beta_2$	$\gamma+\alpha_2+\beta_3$	$\gamma+\alpha_2+\beta_4$	
$\gamma+\alpha_3$	$\gamma+\alpha_3+\beta_2$	$\gamma+\alpha_3+\beta_3$		
$\gamma+\alpha_4$	$\gamma+\alpha_4+\beta_2$			
$\gamma+\alpha_5$				

Adding payment period parameters

- Multiplicative model with parameter for each exposure, development, and **payment** period
- Three "slack factors" mean that we can **drop three parameters** without loss of generality
- Linearize model by applying log transform
- Log-linear model can be represented by parameter vector and model matrix

Adding payment period parameters

- Explore these concepts by studying tabs four to six of "CLRS 2013 - VR1 - Basic Model Structure.xlsx"
- Note that at this point we are not fitting a model to data. We are simply studying the structure of the model.

Reducing number of parameters using trend groups

- Instead of having a "stand-alone" factor for each period, we can also have a model where each factor denotes the change from one period to the next
- Often these trends do not change for multiple periods
- We can exploit this by using one parameter for a group of period to period transitions

Reducing number of parameters using trend groups

- Explore these concepts by studying tabs seven to eight of “CLRS 2013 - VR1 - Basic Model Structure.xlsx”
- Note that at this point we are not fitting a model to data. We are simply studying the structure of the model.

Dealing with missing data points

- The beauty of the matrix representation of the model structure is that excluding a data point from the model simply involves dropping the corresponding row from the model matrix
- We will study how this can be done seamlessly in Excel by exploring the “Each Exp & Dev” tab of “CLRS 2013 - VR1 - Reserves and Ranges.xlsm”

Knowing how many data points are needed for a reserve projection model

- "CLRS 2013 - VR1 - Reserves and Ranges.xlsm" is **not** set up to handle model structures that are not well defined
- If one of the VBA macros goes belly up, don't panic. Close the debugger and turn Automatic Spreadsheet calculation back on.
- Look at the model structure you have specified and try to fix it.

Knowing how many data points are needed for a reserve projection model

- Some things to avoid:
  - Excluding all data points in a row or column
  - Excluding so many data points that the triangle is split up into disconnected regions
  - Grouping parameters in such a way that you end up with a payment period trend group that has no data points in it

Knowing how many data points are needed for a reserve projection model

- A detailed discussion of what can go wrong can be found in my call paper for the 2010 CLRS ("Fitting a GLM to Incomplete Development Triangles")
- We will take some time to explore how "CLRS 2013 - VR1 - Reserves and Ranges.xlsm" reacts when one causes a "crash" and how one recovers from this

Knowing how many data points are needed for a reserve projection model

- "Crashing" aside, try to find the minimum of included data points you need to fit a model and make a reserve projection by specifying such a model in "CLRS 2013 - VR1 - Reserves and Ranges.xlsm"

Issues with incomplete triangles

- Not enough data points for all parameters

γ	$\gamma+\beta_2$	$\gamma+\beta_3$	X	$\gamma+\beta_5$
$\gamma+\alpha_2$	$\gamma+\alpha_2+\beta_2$	$\gamma+\alpha_2+\beta_3$	X	
$\gamma+\alpha_3$	$\gamma+\alpha_3+\beta_2$	$\gamma+\alpha_3+\beta_3$		
$\gamma+\alpha_4$	$\gamma+\alpha_4+\beta_2$			
$\gamma+\alpha_5$				

Issues with incomplete triangles

- Choice of reference cell matters after all

X	X	X	X	X
$\gamma+\alpha_2$	$\gamma+\alpha_2+\beta_2$	$\gamma+\alpha_2+\beta_3$	$\gamma+\alpha_2+\beta_4$	
$\gamma+\alpha_3$	$\gamma+\alpha_3+\beta_2$	$\gamma+\alpha_3+\beta_3$		
$\gamma+\alpha_4$	$\gamma+\alpha_4+\beta_2$			
$\gamma+\alpha_5$				

Issues with incomplete triangles

- Data splits into unrelated regions

X	X	$\gamma+\beta_3$	$\gamma+\beta_4$	$\gamma+\beta_5$
X	X	$\gamma+\alpha_2+\beta_3$	$\gamma+\alpha_2+\beta_4$	
X	X	$\gamma+\alpha_3+\beta_3$		
$\gamma+\alpha_4$	$\gamma+\alpha_4+\beta_2$			
$\gamma+\alpha_5$				

Issues with incomplete triangles

- Exact fit cells

X	X	$\gamma+\beta_3$	$\gamma+\beta_4$	$\underline{\boldsymbol{\gamma+\beta_5}}$
X	X	$\gamma+\alpha_2+\beta_3$	$\gamma+\alpha_2+\beta_4$	
$\gamma+\alpha_3$	$\gamma+\alpha_3+\beta_2$	$\underline{\boldsymbol{\gamma+\alpha_3+\beta_3}}$		
$\gamma+\alpha_4$	$\gamma+\alpha_4+\beta_2$			
$\underline{\boldsymbol{\gamma+\alpha_5}}$				

Part II
Reserve Projections & Ranges

- Learning Objectives
  - Extending loss development methods using GLMs with exposure and development parameters
  - Projecting reserves using GLM's that include payment period parameters
  - Using bootstrapping to simulate the distribution of possible reserve outcomes
  - Interpreting the distribution of reserve outcomes and derived reserve ranges – detecting bias

CLRS 2013 - VR1 - Reserves and Ranges.xlsm

- Input 10 x 10 triangle
- Chose from four different model structures
- Select data points to include in model
- Choose variance function
- Fit GLM to incomplete triangle
- Study standardized residual plots
- Select trend groups (for trend models)
- Bootstrap to get standard error of prediction

Extending LDMs using GLMs with exposure and development parameters

- The same parameters that are used to fit a model to a given triangle, can also be used to predict the values for cells in the bottom triangle of future cells
- Note that the special case of a log-linear GLM with a Poisson distribution function will produce a reserve projection that is identical to the traditional all year volume weighted loss development method

Extending LDMs using GLMs with exposure and development parameters

- Note that the fitted parameters can also be used to fill in the gaps in the triangle left by excluded data points
- We will explore the "Each Exp & Dev" and "Trend Exp & Dev" tabs of "CLRS 2013 - VR1 - Reserves and Ranges.xlsm" to see how this can be done in practice

Projecting reserves using GLM's that include payment period parameters

- While fitting a GLM including payment period parameters works pretty much the same as fitting a GLM with only exposure and development period parameters, there is a big difference for reserve projections:
- Fitting the GLM does not provide us with payment period parameters for future payment periods.

Projecting reserves using GLM's that include payment period parameters

- This is analogous to having to come up with a "tail factor" for development periods
- This is a guess, but we can try to extrapolate trends based on what we see in the triangle
- Note that the scale of payment period parameters has no absolute meaning, because the exposure, development and payment periods are not independent

Projecting reserves using GLM's that include payment period parameters

- We will explore the "Each Exp & Dev & Pay" and "Trend Exp & Dev & Pay" tabs of "CLRS 2013 - VR1 - Reserves and Ranges.xlsm" for one way of projecting future payment period parameters based on the fitted model

Using bootstrapping to simulate the distribution of possible reserve outcomes

- "CLRS 2013 - VR1 - Reserves and Ranges.xlsm" can bootstrap each of the models
- The default behavior is to use linear re-scaling of Pearson residuals, but the application will try to salvage the situation by switching to other re-scaling techniques when necessary
- To avoid "bloating" with simulation data, only summary statistics are saved

What is bootstrapping?

- Approximates the distribution of a function that depends on sampled data
- Assumes that data is randomly distributed according to specified stochastic model
- Uses observed error structure to approximate random distributions of model

Any distributions derived are conditional on specified stochastic model being correct

Bootstrapping and Stochastic Reserving

- Reserves are a function of development triangle
- Get bootstrap distribution of reserve estimates by repeatedly resampling triangle
- Above only gives parameter uncertainty
- To approximate distribution of reserve outcomes we also need process error
- Can approximate process error using the same resampling procedure used for triangle

Bootstrapping and Heteroscedasticity

- Use re-sampling of standardized residuals to adjust for non-constant error structure
- Multiple definitions for residuals available
- Residual re-scaling is the inverse process of residual standardization
- Want to approximate distributions of data points ⇨ <u>resampling distributions should be consistent with stochastic model assumptions</u>

Bootstrapping and Heteroscedasticity

- Linear re-scaling of Pearson residuals can lead to negative values in the resampling distribution (see below) => breakdown of procedure
- The accompanying VBA application tries to salvage this situation by employing split-linear resampling. A separate paper on this technique is currently being written.

Rescaling Example

- Data set Taylor & Ashe (1983)

~~357,848~~	~~766,940~~	~~610,542~~	~~482,940~~	~~527,326~~	~~574,398~~	146,342	139,950	227,229	67,948
~~352,118~~	~~884,021~~	~~933,894~~	~~1,183,289~~	445,745	320,996	527,804	266,172	425,046	
~~290,507~~	~~1,001,799~~	~~926,219~~	1,016,654	750,816	~~146,923~~	495,992	280,405		
~~310,608~~	~~1,108,250~~	776,189	~~1,562,400~~	272,482	352,053	206,286			
~~443,160~~	693,190	991,983	769,488	504,851	470,639				
396,132	937,085	847,498	805,037	705,960					
440,832	847,631	1,131,398	1,063,269						
359,480	1,061,648	1,443,370							
376,686	986,608								
344,014									

Rescaling Example

- Data set Taylor & Ashe (1983) – Fitted Values

140,801	*338,807*	*431,201*	*358,694*	*242,579*	*197,553*	**185,516**	**116,383**	**211,622**	**67,948**
293,186	*705,487*	*897,876*	*746,898*	**505,115**	**411,359**	**386,295**	**242,341**	**440,653**	*141,486*
396,579	*954,279*	*1,214,515*	**1,010,295**	**683,246**	*556,426*	**522,523**	**327,803**	*596,051*	*191,382*
214,098	*515,178*	**655,669**	*545,419*	**368,858**	**300,393**	**282,090**	*176,968*	*321,785*	*103,319*
307,853	**740,778**	**942,791**	**784,261**	**530,383**	**431,937**	*405,619*	*254,464*	*462,697*	*148,564*
343,763	**827,188**	**1,052,766**	**875,744**	**592,251**	*482,321*	*452,933*	*284,146*	*516,669*	*165,893*
386,316	**929,583**	**1,183,083**	**984,148**	*665,564*	*542,025*	*509,000*	*319,320*	*580,625*	*186,429*
442,821	**1,065,549**	**1,356,128**	*1,128,096*	*762,913*	*621,305*	*583,450*	*366,025*	*665,551*	*213,697*
400,230	**963,064**	*1,225,695*	*1,019,595*	*689,536*	*561,548*	*527,333*	*330,821*	*601,538*	*193,143*
344,014	*827,792*	*1,053,534*	*876,383*	*592,684*	*482,673*	*453,264*	*284,354*	*517,046*	*166,014*

Rescaling Pearson Residuals

- Definition residuals:

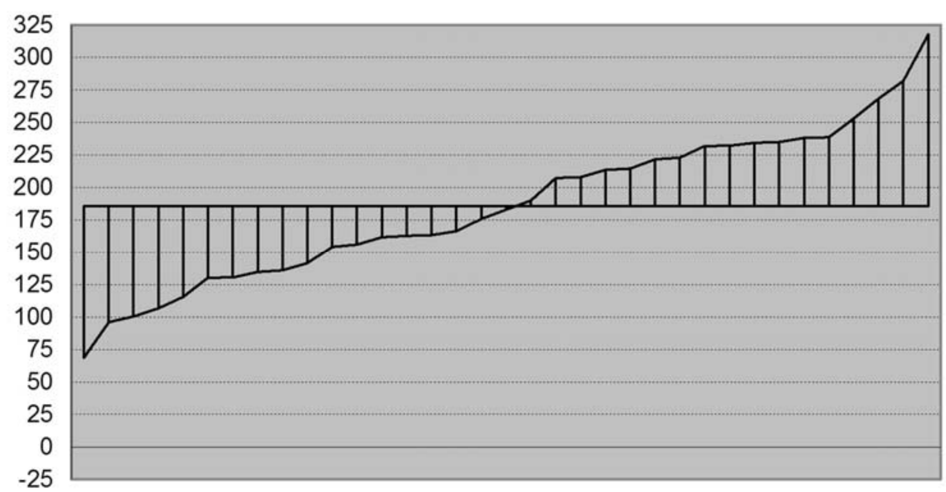
$$r_P = \frac{y - \hat{y}}{\sqrt{V(\hat{y})}}$$

- Definition resampling distribution:

$$y_P^* = \hat{y} + \sqrt{V(\hat{y})} \cdot \boldsymbol{S}$$

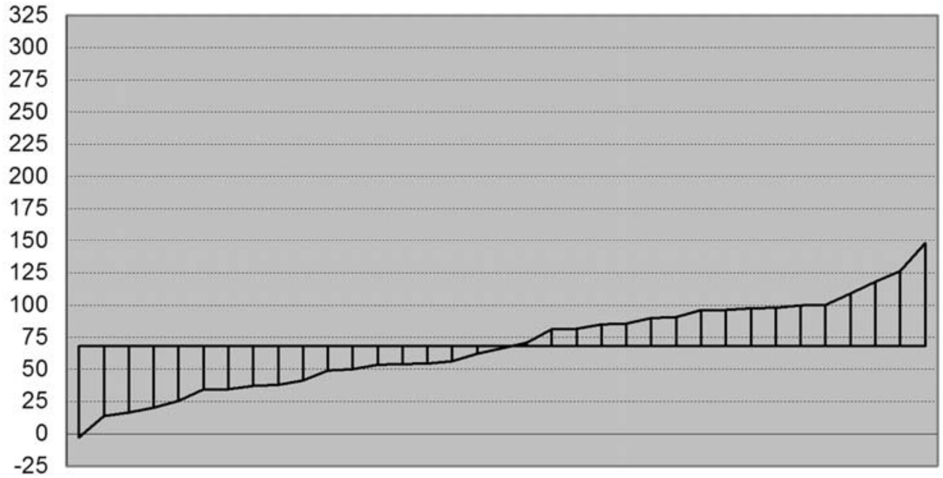
Rescaling Pearson Residuals

- Resampling distribution - fitted mean of 185,586



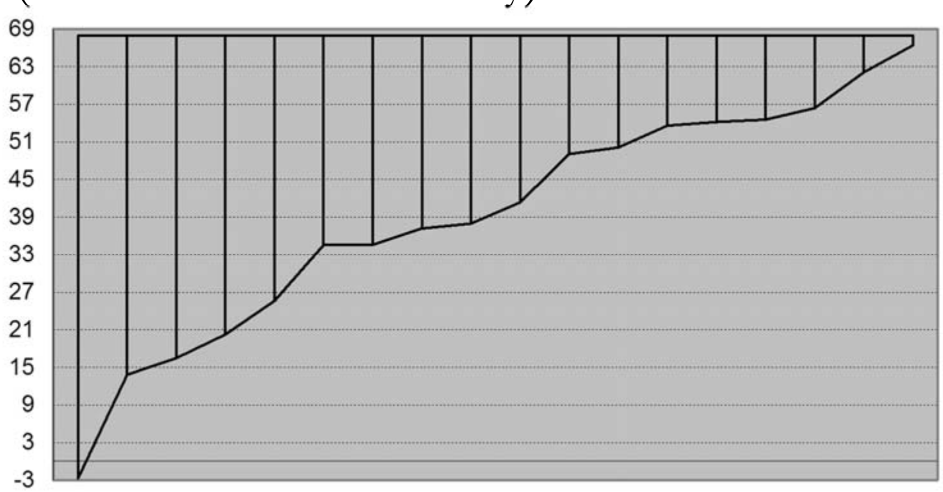
Rescaling Pearson Residuals

- Resampling distribution - fitted mean of 67,948



Rescaling Pearson Residuals

- Resampling distribution - fitted mean of 67,948 (values below mean only)



Projecting reserves using GLM's that include payment period parameters

- Note how accompanying spreadsheet application can bootstrap the above model without any problems.
- A detailed discussion of what can go wrong with re-scaling can be found in my call paper for the 2010 CLRS ("Bootstrapping Generalized Linear Models for Development Triangles Using Deviance Residuals ")

Interpreting the distribution of reserve outcomes and derived reserve ranges

- Bootstrapping provides a distribution of possible reserve outcomes that is conditional on the underlying stochastic model being correct
- The same assumption underlies confidence interval calculations in classical statistics
- Bootstrapping does not account for model error
- Bootstrapping can reveal bias that is inherent in the projection method

Interpreting the distribution of reserve outcomes and derived reserve ranges

- Examine the bootstrapping results for the four different models:
  - Do you detect any bias?
  - How many bootstrap iterations seem necessary to distinguish bias from random fluctuations

Contact Information

- Accompanying spreadsheets can be downloaded from the 2013 CLRS web-site
- For any questions please contact Thomas Hartl

- thartl@bryant.edu
- 401-232-6124