Bootstrapping Generalized Linear Models for Development Triangles Using Deviance Residuals

Thomas Hartl, ACAS

Abstract: This paper presents a practical study of how to bootstrap a development triangle using a generalized linear model (GLM) and deviance residuals. We also point out some limitations inherent in bootstrapping approaches. (Interested readers can contact the author and request a copy of an MS Excel application to further explore the concepts discussed in this paper.) First we demonstrate how Pearson residual bootstrapping can fail when applied to GLMs because of their linear rescaling properties. Next we describe an algorithm for rescaling deviance residuals based on the identity variance function. We continue with an example where Pearson residual bootstrapping fails, while deviance residuals bootstrapping works. We then present bootstrap simulation results for two GLMs: one where both approaches work and the original example where only deviance residuals can be applied. Subsequently we prove that deviance residuals based on the identity variance function are bounded below for any given data point with the lower bound depending on the fitted value for the data point. We then give an example of a GLM where deviance residual bootstrapping fails because of this property. The paper concludes with a discussion "distribution-free" versus parametric resampling.

Keywords. Bootstrapping and Resampling Methods, Generalized Linear Modeling, Reserve Variability, Reserving Methods, Nonparametric Methods.

1. INTRODUCTION

In the context of stochastic reserving, several authors (e.g., [4], [7] and [9]) have stressed the need of casting the task of projecting reserves in a rigorous way as a regression problem. Several authors (e.g., [4], [7] and [9]) have also pointed out that performing an all-years, volume weighted, link-ratio estimate leads to the same result as fitting a GLM with the logarithmic link function and the identity variance function. Some papers (e.g., [4]) and many practitioners have exploited this equivalence to implement spreadsheet based applications for deriving a distribution of possible reserve outcomes based on bootstrap simulations by repeated resampling and application of the link-ratio estimate. While suitable to illustrate the concept of bootstrapping, these applications are typically not flexible enough to deal with practical judgments reserving analysts have to make about which cells of the triangle are deemed to be representative of future development (e.g., use data from last *n* diagonals or exclude obvious abnormalities). In [5] the author of this paper presented a practical account of how to rigorously translate such judgments into a well-defined regression problem using the apparatus of GLM theory. This paper builds on the framework established in [5] and presents a case study to give a practical demonstration of some limitations inherent in non-parametric bootstrapping approaches based on Pearson or deviance residuals.

1.1 Research Context

General accounts of how to apply bootstrapping methods to a GLM for a incremental development triangle have been provided in [4] and [7]. Details on how to apply bootstrapping to a GLM can also be found in chapter 7.2 of [3]. This paper demonstrates that the common Boostrapping approach based on Pearson residuals does break down when the linear rescaling of residuals leads to negative incremental values in the resampling distribution for some triangle cells. In [3], [4], and [7] the authors do mention that there are also alternative ways of defining residuals. We illustrate that a bootstrapping approach based on deviance residuals (using the identity variance function) may succeed where bootstrapping based on Pearson residuals fails. As we demonstrate, this alternative approach also has practical limitations. At least for the identity variance function, deviance residuals are technically not identically distributed, and it may not be possible to rescale all residuals for resampling purposes. Throughout this paper we use the GLM framework for incomplete development triangles established in [5]. Our case study suggests that there may be good practical reasons to prefer parametric resampling over nonparametric resampling.

1.2 Objective

Bootstrapping has become a popular method for deriving distributions of reserve outcomes based on development triangles. This paper provides a case study to demonstrate some inherent limitations of applying versions of this method to a GLM for an incremental development triangle. These practical limitations point to the need for further research into alternative resampling schemes. We also hope that the case study and the discussion of the issues encountered will provide readers with a better understanding of what bootstrapping really is.

1.3 Outline

The remainder of the paper proceeds as follows. Section 2 explains the difference between linear and non-linear rescaling of residuals, introduces a version of a Newton-Raphson algorithm for rescaling deviance residuals based on the identity variance function, and presents an example of a GLM for an incremental development triangle for which bootstrapping with deviance residuals is possible while bootstrapping with Pearson residuals fails. In section 3 we analyze a limitation of bootstrapping with deviance residuals based on the identity variance function. We demonstrate that the theoretical distribution of deviance residuals for a given triangle cell is bounded below, and that the bound varies with the square root of the expected mean. This means that negative deviance residuals for triangle cells with larger expected means may be "out of bounds" for some triangle cells with smaller expected means. We provide an example of a GLM for an incremental development triangle where bootstrapping with deviance residuals is not possible for this reason. In the summary and discussion section we review our results provide and invite the reader to explore the concepts presented with the accompanying MS Excel file. In the conclusion we reflect on the "distribution free" or "non-parametric" attributes of bootstrapping approaches. We suggest that in the context of GLMs (or other stochastic models) for development triangles "parametric" resampling may be just as useful while avoiding some of the limitations demonstrated in this paper.

2 RESCALING OF RESIDUALS FOR RESAMPLING

In general bootstrapping deals with the heteroscedasticity of the underlying stochastic model by rescaling the residuals obtained from a specific data point so they can be applied to the expected means of other data points. The procedure for rescaling depends on the definition of residual used. While various residuals are mentioned in [4] and [7], the examples presented in these papers use Pearson residuals. Here we demonstrate a limitation imposed by the linear rescaling of Pearson residuals and contrast this with the non-linear rescaling properties of deviance residuals.

2.1 Linear vs. non-linear rescaling

We assume that we have a collection of suitably standardized residuals that were obtained by fitting a GLM to an incremental development triangle using pseudo-likelihood with the identity variance function and the natural logarithm as the link function. Suitably standardized means that the standardized residuals for various triangle cells can be considered as being approximately independent identically distributed (iid). We will return to the question of whether deviance residuals can be considered iid in section 3 of this paper. For the time being we simply follow the standardization suggested by equations 12.4 and 12.5 on page 397 in [6]. So, in continuing we assume that we a have vector of standardized residuals denoted by $\mathbf{s}$. Note that technically this vector will be defined differently for Pearson and deviance residuals, but we will not distinguish this in our notation. For any given data point we can think of the residuals as a measure of how much an actual observation differs from the expected value for that observation. If the expected value, $\hat{\mathrm{y}}$, for the observation is considered a constant parameter the residual, r, is a function of the actual observation, y, alone:

$$r = \mathrm{H}_{\hat{\mathrm{y}}}\left(y\right). \tag{2.1}$$

In abstract terms the rescaling of the standardized residuals, **s**, is accomplished by applying the functional inverse of $\mathrm{H}_{\hat{y}}$ to the elements of **s**. So the vector of resampling values **y*** can be defined by:

$$\mathbf{y}^* = \mathrm{H}_{\hat{y}}^{-1}(\mathbf{s}). \tag{2.2}$$

Noting that $\mathrm{V}(\hat{y})$ stands for the variance function of the expected mean of a data point, equation (2.1) for Pearson residuals becomes:

$$r_{\mathrm{P}} = \frac{y - \hat{y}}{\sqrt{\mathrm{V}(\hat{y})}}. \tag{2.3}$$

As one can seem this is a linear function of y, and equation (2.2) therefore also takes linear form:

$$\mathbf{y}_{\mathrm{P}}^* = \hat{y} + \sqrt{\mathrm{V}(\hat{y})} \cdot \mathbf{s}. \tag{2.4}$$

So the distribution of resampling values is a linear transformation of the standardized residuals, with the same rescaling factor being applied to all standardized residuals.

Since the standardized residuals have a mean of zero,[1] this also means that for sufficiently small expected values, $\hat{y}$, there will always be negative values in the resampling distribution. For a GLM with a logarithmic link function, this represents a violation of the fundamental model assumption of positive incremental values and the MLE algorithm cannot be applied since we cannot take the logarithm of a negative number.

In the context of quasi-likelihood estimation, equation (2.1) for deviance residuals can generally be expressed as follows (see equation 9.4 on page 327 in [6]):

$$r_{\mathrm{D}} = \mathrm{sign}(y - \hat{y}) \cdot \sqrt{2\int_{\hat{y}}^{y} \frac{y - t}{\mathrm{V}(t)} dt}. \tag{2.5}$$

Needless to say that dealing with this sort of expression is mathematically more complex than the relatively simple functional form of equation (2.3). In particular it is not possible to give a general expression for the functional inverse analogous to equation (2.4). We will explain how to approach this task numerically for the identity variance function in the next subsection. We conclude this subsection by noting that while highly non-linear, deviance residuals are perfectly well-behaved, and provided that $\mathrm{V}(t)$ goes to 0 sufficiently fast as t goes to 0, we are guaranteed that inverting the function will not result in negative resampling values.

[1] If necessary we will make a centering adjustment to guarantee that this is the case.

2.2 Rescaling deviance residuals based on the identity variance function

With the identity variance function (i.e., V(ŷ)= ŷ) the integral in equation (2.6) can be given in closed form and we get the following expression (see expression for Poisson distribution on page 39 in [6]) for the deviance residuals:

$$r_{\mathrm{D}} = \operatorname{sign}(y-\hat{y}) \cdot \sqrt{2\left(y \cdot \log\left(y/\hat{y}\right) - y + \hat{y}\right)}\,. \tag{2.6}$$

We are still not in a position to give a closed-form expression for the functional inverse, but we can numerically solve this using a variant of the Newton-Raphson algorithm (combined with the bisection method) based on the code provided on pages 366/7 in [8].[2] Since this is a well documented standard algorithm we will not go into all implementation details here. We will, however, give some details on the modifications we have made to tweak this to the concrete task at hand. Firstly, to simplify treatment of the sign of the residual and to get rid of the square root, we actually just invert the right-hand side of the following equation:

$$\left(r_{\mathrm{D}}\right)^2 = 2 \cdot \hat{y} \cdot \left(y/\hat{y} \cdot \log\left(y/\hat{y}\right) - y/\hat{y} + 1\right). \tag{2.7}$$

Note that the right-hand side of this equation does not define a one-on-one function, so we need to choose an appropriate domain (i.e., upper and lower bound for y) based on the sign of the residual in question. To further simplify, we substitute $x = y/\hat{y}$ and w = $(r_{\mathrm{D}})^2 / (2 \cdot \hat{y}) - 1$, and thus arrive at:

$$w = x \cdot \log\left(x\right) - x\,. \tag{2.8}$$

In order to numerically solve this for x, we need upper and lower bounds depending on the value of w. With negative residuals, it suffices to restrict x to (0,1). For positive residuals the lower bound for x is clearly 1, but our algorithm uses more refined initial estimates as detailed in appendix 1. With this set-up, we solve for x in terms of w by using the Newton-Raphson algorithm to find the zero of the function $f(x)$ defined by:

$$f(x) = x \cdot \log\left(x\right) - x - w\,. \tag{2.9}$$

For those readers not familiar with the Newton-Raphson method, the algorithm proceeds by iterating over x until convergence is achieved, using the following formula:

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}\,. \tag{2.10}$$

[2] A pdf of the cited section is freely available at http://www.nrbook.com/a/bookcpdf/c9-4.pdf.

For completeness, in our case the derivative of $f(x)$ is given by:

$$f'(x) = \log(x). \tag{2.11}$$

As indicated above the actual algorithm implemented is mixing the Newton-Raphson method with the bisection method to prevent x from jumping out of bounds or to improve the speed of convergence when $f(x) << f'(x)$.

The interested reader can contact the author and request a copy of the companion MS Excel application to study and explore the source code of user defined function VB_PoissonDevianceResidual_Inverse. Also note that, as implemented here, the algorithm assumes that residuals have not been adjusted (i.e., normalized) for the dispersion factor.

Now that we know how to compute resampling distributions using both Pearson and deviance residuals, we can apply this apparatus to real-life data.

2.3 Example

We are using a data set that the authors of [7] attribute to Taylor and Ashe (1983). Here is the data in incremental form:

~~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									

To avoid visual clutter we have omitted row and column labels. There are no non-positive or missing data points, but we have chosen only to include the latest five diagonals of incremental values and to exclude three further triangle cells that were identified as outliers in preliminary analysis. Excluded data points from the original data set have been indicated by "~~strikethrough~~" formatting.

Fitting a GLM with the logarithm as link function and the identity variance function results in the following fitted values:

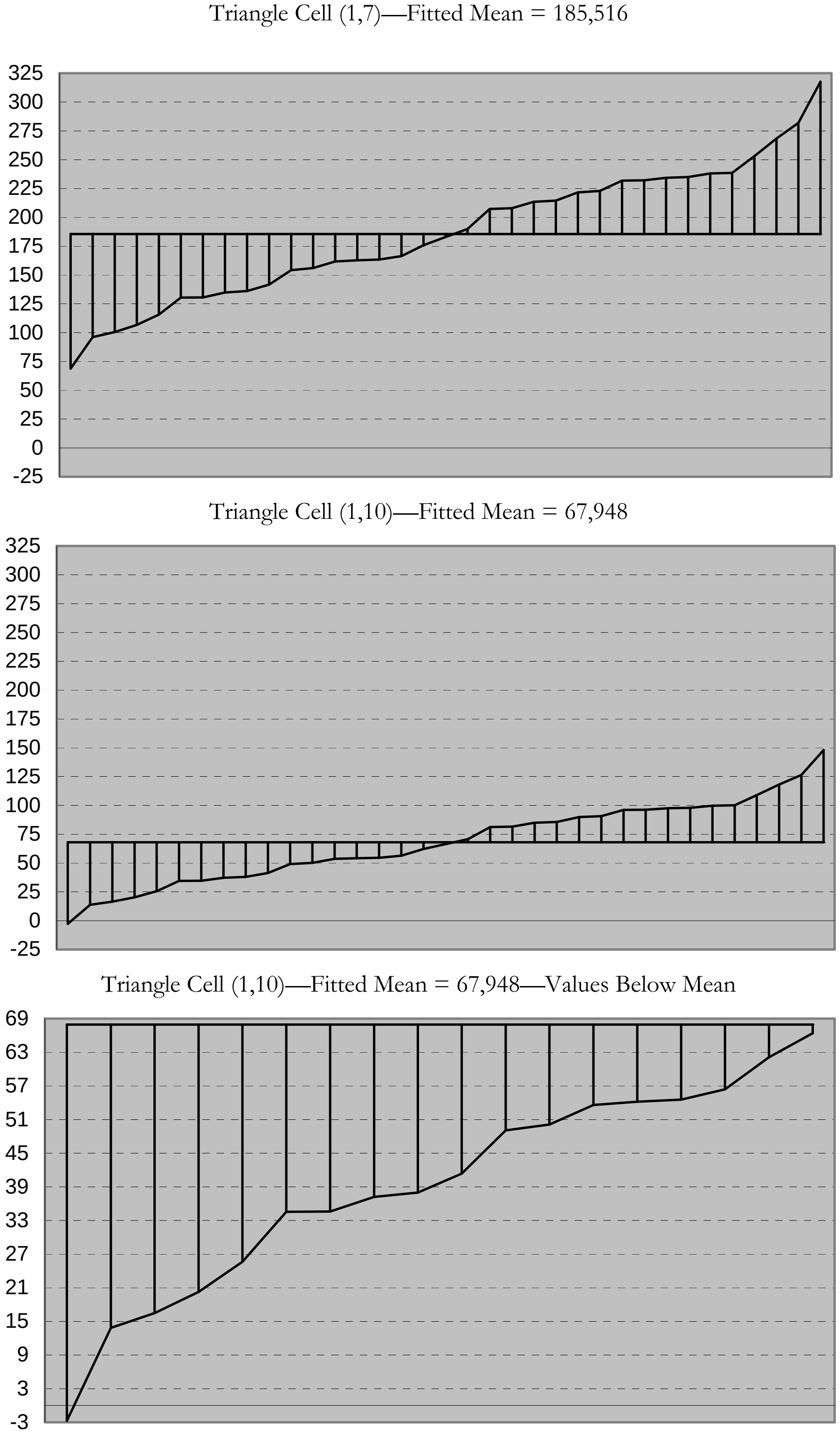
140,801	*338,807*	*431,201*	*358,694*	*242,579*	*197,553*	**185,516**	**116,383**	**211,622**	**67,948**
293,186	*705,487*	*897,876*	*746,898*	**505,115**	**411,359**	**386,295**	**242,341**	**440,653**	*141,486*
396,579	*954,279*	*1,214,515*	**1,010,295**	**683,246**	*556,426*	**522,523**	**327,803**	*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*

As the author of this paper has demonstrated in [5], this type of GLM does not only project expected values for future triangle cells but also extrapolates the expected values for all past triangle cells that were excluded from the analysis. In the above table, we show all fitted values that correspond to included data points in bold letters. All values in italics correspond to projections/extrapolations based on the fitted parameters for the model. Since we have chosen the identity variance function the reader can also verify that fitted values in bold preserve the row and column sums of the original data points for the included triangles cells.

To bootstrap this GLM, we repeatedly generate pseudo-data for each of the included triangle cells (see bold-face fitted values in the above table). During each iteration step we re-estimate the GLM based on the pseudo-data generated. For a stochastic reserving application we also calculate the estimated reserve and save the total by accident year or in aggregate to get a simulated distribution of reserve estimates to evaluate the inherent parameter error. Typically one would similarly simulate future development amounts to account for the process error.

Effectively this resampling process defines a resampling distribution for each data point, which is obtained by rescaling all available standardized residuals and applying them to the fitted values as described in the previous two subsections. These resampling distributions can be pre-computed and stored both to save execution time during each bootstrap iteration, and to evaluate them for consistency with the underlying model assumptions. Here our main concern is with negative incremental values, since these prevent us from fitting the model with the MLE algorithm. On the following two pages we graph the resampling distributions for two of the triangle cells. Figure 1 shows the results obtained based on Pearson residuals and Figure 2 shows the corresponding results using deviance residuals.

Resampling Distributions with Pearson Residuals

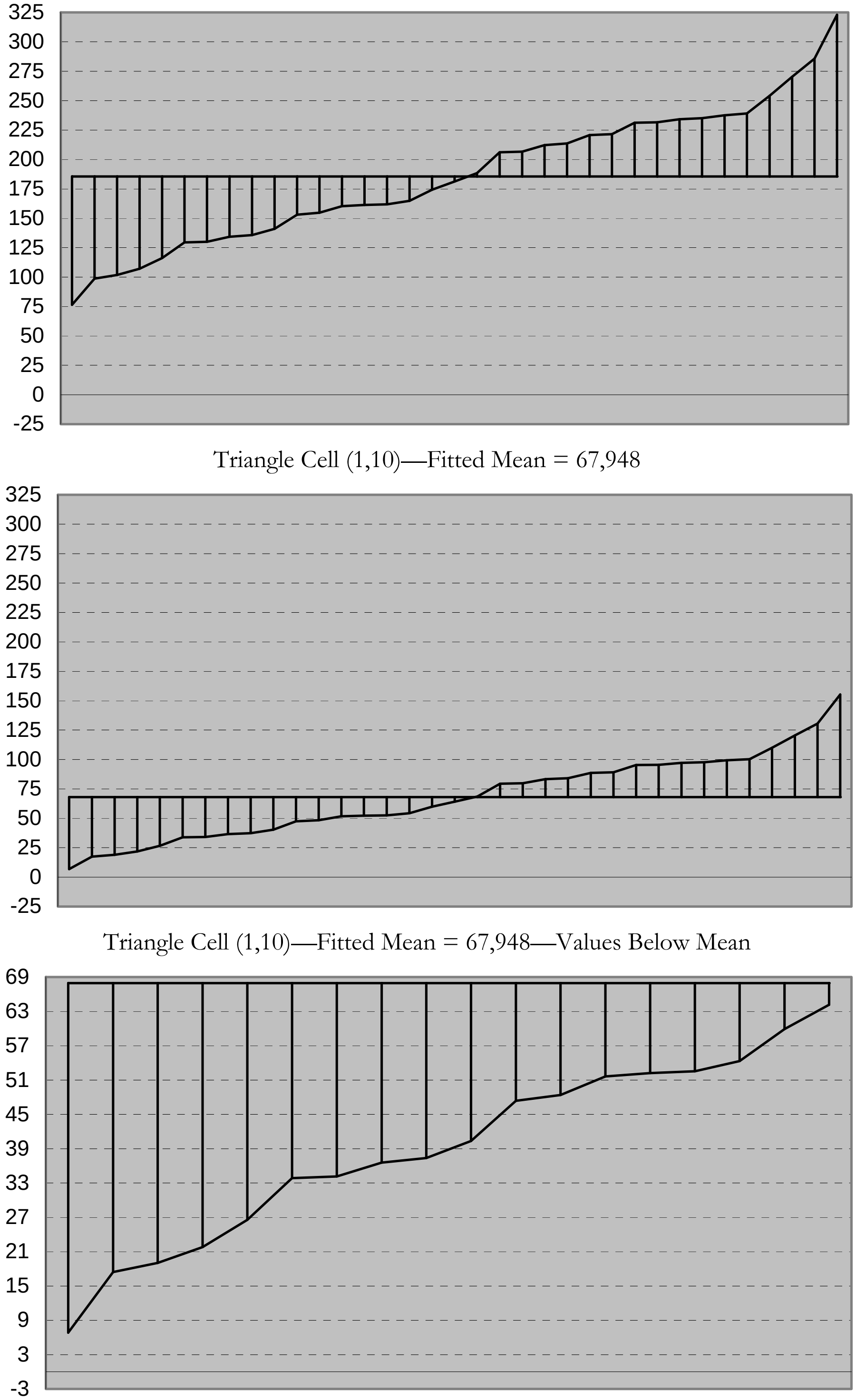


Triangle Cell (1,7)—Fitted Mean = 185,516

Triangle Cell (1,10)—Fitted Mean = 67,948

Triangle Cell (1,10)—Fitted Mean = 67,948—Values Below Mean

Figure 1



Resampling Distributions with Deviance Residuals

Triangle Cell (1,7)—Fitted Mean = 185,516

Triangle Cell (1,10)—Fitted Mean = 67,948

Triangle Cell (1,10)—Fitted Mean = 67,948—Values Below Mean

Figure 2

Overall the two sets of resampling distributions are very similar. Comparing the top graphs for triangle cell (1,7) one can see that the smallest value obtained by Pearson residual resampling is below that obtained by deviance residual resampling. The same holds true for the largest value. The fact that both distributions nevertheless have the same mean of 185,516 (and approximately the same variance) is the first indication that deviance residuals are subtly different.

Comparing the middle graphs for triangle cell (1,10) to the corresponding top graphs for cell (1,7) one notices how the spread from minimum to maximum resampling value is smaller for triangle cell (1,7) than for triangle cell (1,10). This is the effect of the rescaling to ensure that each resampling distribution has the appropriate variance as defined by the variance function assumed in the GLM specification.

Since our main concern here is with negative incremental values we have included slightly larger scale graphs of the lower part of the resampling distribution for triangle cell (1,10) at the bottom of figure 1 and 2. As we can see, Pearson resampling leads to one negative resampling value, while deviance resampling stays positive for the corresponding residual.

In this particular example the negative resampling value for cell (1,10) is actually the only negative value resulting from Pearson resampling. Nevertheless it does mean that the GLM as specified cannot be bootstrapped using Pearson residuals. One alternative would be to exclude cell (1,10) from the model specification. In this particular case this leads to a model that projects a reserve of versus a reserve of for the original model including cell (1,10). Noting that excluding cell (1,10) amounts to assuming zero development for the 10th development period, it is not surprising that the projected reserve is smaller.

To demonstrate the differences in outcomes, we conclude this section with output obtained by bootstrapping the model excluding cell (1,10) both with Pearson and with deviance residuals. We also include the bootstrapping results for the model including cell (1,10). The "Modeled Reserve" column of each output table shows the reserve projection representing the expected future development amounts that result from fitting the GLM. The "Bootstrap Projection" column shows the mean of the reserve projections based on simulated data. Note that the "Bootstrap Projection" distribution defines the parameter error. The "Simulated Future Development" column shows the mean of the simulated future development amounts used to incorporate process error. Comparing the "Modeled Reserve" to "Simulated Future Development" also allows one to gauge whether we have sufficiently many bootstrap iterations to keep bias resulting from sampling error to an acceptable level. The "Standard Prediction Error" is the root of the mean square error of simulated reserve outcomes (i.e., projected reserve based on pseudo-data less simulated future development). All the way to the right we also show a confidence interval based on empirical percentiles of simulated reserve outcomes. Note that a positive number represents a reserve projection above the simulated future development amount.

Bootstrapping Results with 10,000 Iterations

Excluding Cell (1,10)—Pearson Residuals

Accident Period	Modeled Reserve	Bootstrap Projection	Sim. Future Development	Standard Pred. Error	5%-ile Sim. Outcome	95%-ile Sim. Outcome
1	-	-	-	-	-	-
2	-	-	-	-	-	-
3	596,051	603,398	595,127	166,522	(254,940)	288,038
4	498,753	504,064	498,789	135,273	(214,047)	231,751
5	1,122,779	1,134,746	1,125,780	224,917	(345,901)	394,656
6	1,736,070	1,751,181	1,734,825	302,852	(467,485)	522,686
7	2,616,534	2,640,194	2,612,849	407,758	(613,245)	724,636
8	4,127,340	4,164,901	4,132,367	586,633	(892,087)	1,040,074
9	4,956,065	4,990,267	4,959,138	801,618	(1,232,452)	1,417,929
10	5,087,731	5,161,854	5,082,052	1,393,141	(2,030,612)	2,510,119
Total	**20,741,324**	**20,950,606**	**20,740,927**	**2,504,915**	**(3,645,668)**	**4,603,584**

Excluding Cell (1,10)—Deviance Residuals

Accident Period	Modeled Reserve	Bootstrap Projection	Sim. Future Development	Standard Pred. Error	5%-ile Sim. Outcome	95%-ile Sim. Outcome
1	-	-	-	-	-	-
2	-	-	-	-	-	-
3	596,051	601,425	595,682	165,133	(254,824)	283,543
4	498,753	502,686	497,483	135,937	(213,619)	233,681
5	1,122,779	1,130,897	1,122,776	225,374	(348,761)	388,058
6	1,736,070	1,748,560	1,735,691	300,235	(460,344)	514,240
7	2,616,534	2,636,940	2,619,302	409,205	(630,112)	709,919
8	4,127,340	4,156,304	4,128,423	582,196	(885,958)	1,016,114
9	4,956,065	5,002,022	4,962,549	802,422	(1,215,009)	1,405,855
10	5,087,731	5,169,300	5,088,560	1,404,841	(2,048,259)	2,501,894
Total	**20,741,324**	**20,948,135**	**20,750,465**	**2,530,813**	**(3,764,693)**	**4,599,894**

Including Cell (1,10)—Deviance Residuals

Accident Period	Modeled Reserve	Bootstrap Projection	Sim. Future Development	Standard Pred. Error	5%-ile Sim. Outcome	95%-ile Sim. Outcome
1	-	-	-	-	-	-
2	141,486	148,558	141,810	99,435	(142,027)	181,427
3	787,433	802,512	786,345	227,758	(332,132)	415,547
4	602,073	612,774	600,459	168,556	(252,157)	302,197
5	1,271,343	1,290,547	1,271,004	266,900	(394,291)	476,089
6	1,901,963	1,926,750	1,906,391	343,783	(513,444)	607,984
7	2,802,963	2,834,990	2,804,315	448,446	(679,871)	795,858
8	4,341,037	4,384,089	4,338,730	639,559	(958,332)	1,144,621
9	5,149,209	5,209,231	5,145,549	844,468	(1,259,637)	1,509,566
10	5,253,745	5,354,869	5,249,988	1,444,013	(2,074,331)	2,567,191
Total	**22,251,251**	**22,564,319**	**22,244,592**	**2,868,629**	**(4,054,094)**	**5,235,817**

Figure 3

3 A PRACTICAL LIMIT OF DEVIANCE RESIDUALS

In the previous section we demonstrated that the non-linear rescaling properties of deviance residuals allow us to bootstrap a GLM in some instances where Pearson residuals lead to negative values in the resampling distribution for some data points. This does not mean that any GLM for an incomplete development triangle can be bootstrapped using deviance residuals. In this brief section we explore the mathematical reason for why this is the case for deviance residuals based on the identity variance function. We also give an example of a GLM based on the same data set used in the previous section where deviance residual resampling cannot be applied.

3.1 Taking the limit

For convenience we repeat the definition of deviance residuals based on the identity variance function (i.e., equation 2.):

$$r_{\mathrm{D}} = \mathrm{sign}(y-\hat{y}) \cdot \sqrt{2\left(y \cdot \log\left(y/\hat{y}\right) - y + \hat{y}\right)}\,. \tag{3.1}$$

Here we are interested in the lower limit as $y \rightarrow 0$, hence we can substitute -1 for sign(y-ŷ). After some rearranging we obtain:

$$\lim_{y\to 0} r_{\mathrm{D}} = -\sqrt{2 \cdot \lim_{y\to 0}\left(\frac{\log\left(y/\hat{y}\right)}{y^{-1}} - y + \hat{y}\right)}\,. \tag{3.2}$$

Dealing with the easy parts we get:

$$\lim_{y\to 0} r_{\mathrm{D}} = -\sqrt{2} \cdot \sqrt{\lim_{y\to 0}\left(\frac{\log\left(y/\hat{y}\right)}{y^{-1}}\right) + \hat{y}}\,. \tag{3.3}$$

The remaining part goes to -∞/∞ as $y \rightarrow 0$, so we can use l'Hôpital's rule to evaluate it, leading to:

$$\lim_{y\to 0} r_{\mathrm{D}} = -\sqrt{2} \cdot \sqrt{\lim_{y\to 0}\left(\frac{\left(y/\hat{y}\right)^{-1}}{-y^{-2}}\right) + \hat{y}}\,. \tag{3.4}$$

This simplifies to:

$$\lim_{y\to 0} r_{\mathrm{D}} = -\sqrt{2} \cdot \sqrt{\lim_{y\to 0}\left(-y \cdot \hat{y}\right) + \hat{y}} = -\sqrt{2\hat{y}}\,. \tag{3.5}$$

So we can see that for any given triangle cell the smallest theoretical value for the deviance residual is $-(2\hat{y})^{.5}$. Obviously this result is dependent on the particular functional form of the deviance residual, which in the case of the identity variance function is given by equation 3.1.

This does raise the question of whether deviance residuals can be considered approximately iid, which is a fundamental underlying assumption of resampling methods. Theory aside, we are left with a practical issue when trying to use deviance residuals for bootstrapping: a deviance residual obtained from a data point with a larger expected mean may be below the lower bound for deviance residuals for a data point with a smaller expected mean. If this happens we cannot rescale the deviance residual in question for the purpose of resampling. So, if $\hat{y}_{\min}$ is the smallest fitted value for a particular GLM with the identity variance function, we can only use deviance residual bootstrapping, if for all data points included in the model we have

$$r_{\mathrm{D}} > -\sqrt{2\cdot\hat{y}_{\min}}\ . \tag{3.6}$$

We will now demonstrate that it is not difficult to come up with an example where this relationship does not hold for all data points.

3.2 Example

The GLM used as an example here is not very different from the one introduced in section 2.3. We include cells (1,6) and (3,6) which were previously excluded:

~~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									

As before we fit a GLM with a logarithmic link function and the identity variance function. This results in the following fitted values:

254,672	*611,704*	*774,193*	*665,389*	*434,726*	**320,588**	**299,529**	**184,715**	**283,087**	**67,948**
332,131	*797,756*	*1,009,667*	*867,770*	**566,950**	**418,096**	**390,632**	**240,897**	**369,188**	*88,615*
359,730	*864,049*	*1,093,569*	**939,880**	**614,062**	**452,839**	**423,093**	**260,915**	*399,867*	*95,978*
223,757	*537,449*	**680,214**	*584,618*	**381,955**	**281,672**	**263,169**	*162,293*	*248,723*	*59,700*
311,253	**747,608**	**946,198**	**813,221**	**531,310**	**391,814**	*366,076*	*225,754*	*345,981*	*83,044*
343,043	**823,968**	**1,042,841**	**896,282**	**585,577**	*431,833*	*403,467*	*248,812*	*381,319*	*91,526*
384,679	**923,974**	**1,169,412**	**1,005,065**	*656,650*	*484,245*	*452,436*	*279,011*	*427,600*	*102,635*
444,666	**1,068,059**	**1,351,772**	*1,161,796*	*759,049*	*559,759*	*522,990*	*322,520*	*494,280*	*118,640*
400,741	**962,553**	*1,218,240*	*1,047,030*	*684,067*	*504,464*	*471,327*	*290,660*	*445,454*	*106,920*
344,014	*826,299*	*1,045,792*	*898,818*	*587,234*	*433,055*	*404,608*	*249,516*	*382,397*	*91,785*

We can see that if $\hat{y}_{\min}$ = 67,948 from cell (1,10). For bootstrapping purposes this results in a lower bound of -368.64 for (unscaled) deviance residuals. Applying equation 3.1 to cell (3,6) with y = 146,923 and $\hat{y}$ = 452,839 we get the following (unscaled) deviance residual:

$$r_{\mathrm{D}} = -530.16\,. \tag{3.7}$$

Hence we can see that it is not possible to bootstrap this GLM using deviance residuals. Finally, our discussion has focused on the resampling of included data points. If the residuals are also used for simulating the process error, the allowable minimum would also depend on the smallest expected value for future development periods.

4. RESULTS AND DISCUSSION

In section 2.1 we showed how resampling with Pearson residuals can lead to negative incremental values in the resampling distribution. This in turn means that we cannot apply the MLE algorithm to for fitting the GLM during bootstrap iterations. We also presented the concept of deviance residuals and explained how these hold the promise of avoiding the issue of negative incremental values.

Section 2.2 provided details on how to compute the inverse of the deviance residual function based on the identity variance function. The algorithm is based on a variation of the Newton-Raphson method.

In section 2.3 an example of a GLM was presented, where bootstrapping with Pearson residuals is not possible, but bootstrapping with deviance residuals works. The crucial difference in the resampling distributions resulting from the different techniques was graphically illustrated. We also showed bootstrapping output for a slightly modified model where both approaches work and for the main example where only the deviance residual approval can be applied. It should be noted that deviance residuals therefore can broaden the scope of bootstrapping approaches.

In section 3.1 we proved that for any given data point, the deviance residuals for log-link GLMs with the identity variance function are bounded below $-(2\hat{y})^{.5}$, where $\hat{y}$ is the fitted value for that data point. This result means that Boostrapping with deviance residuals can only work if all standardized deviance residuals exceed the lower bound for the smallest fitted value. In section 3.2 we presented another GLM (again a slightly modified version of the example in section 2.3) where this condition is indeed violated.

The material in this paper is based on standard GLM theory and standard numerical methods. We hope this paper contributes to making more actuaries aware of how these powerful methods can practically be applied in the context of stochastic reserving. Given the popularity of bootstrapping approaches, we also feel that it is also important to draw attention to some of their inherent limitations.

Interested readers are encouraged to contact the author and request a copy of the companion MS Excel application to further explore the concepts and algorithms presented in this paper. Interacting with this application should prove a useful aid to gaining a deeper understanding of what regression models can accomplish in the context of development triangles. As far as the bootstrapping functionality is concerned, the application follows the approach outlined in [7]. In particular we use "procedure 2" as outlined in Figure 2 of that paper. As pointed out in [7], this is also the approach described in [3].

We want to conclude this section reflecting on "distribution-free" (or "non-parametric") versus parametric approaches to bootstrapping. Before bootstrapping was applied in an actuarial context, it was introduced as stochastic modeling techniques for detecting bias for estimators or to derive confidence intervals for parameter estimates in cases where standard regression may not work well because the underlying error structure is not normally distributed (or does not follow other known error distributions, for which specialized regression techniques are available). Especially in cases where there is a decent number of observations which all can be assumed to come from the same underlying distribution, bootstrapping can provide results that are superior to those obtained by applying standard regression techniques, based on assumptions not satisfied by the data at hand. This is the context in which "distribution-free" approaches shine.

The question to consider here is "What makes bootstrapping attractive in the context of stochastic reserving?" One advantage is that bootstrapping can derive a distribution for just about any function one may want to calculate based on an observed sample of data points. Note that the estimated reserves technically are not the fitted model itself, but a projection of expected amounts for future (i.e., out of sample) development periods. Hence bootstrapping is useful, because we can just recalculate the reserve based on the pseudo data generated for each iteration, and thus we simulate an empirical distribution of reserve estimates. There is no need to theoretically understand how the errors in the parameter estimates of the underlying model (and their correlations) might affect the distribution of reserve estimates.

The "distribution free" aspect of bootstrapping does not appear to be particularly important in the context of stochastic reserving based on development triangles. Often the data sample (number of triangle cells) is not particularly large in relation to the number of parameters we are trying to estimate. Furthermore, it is not obvious that the error structure for incremental development in the first year, for example, is in anyway systematically related to the error structure in the fifth year. Variation in the first year may be due to subtle variation in the mix of claims for subcoverages, while fifth-year development might be caused by sporadic late reporting claims, or an unexpected judicial decision for a single open claim. Non-parametric bootstrapping is based on the assumption that (after standardization and dealing with heteroscedasticity) the residuals are the best available approximation for the error structure driving the underling stochastic process.

So, can we harvest the power of bootstrapping as a simulation technique for deriving a distribution of reserve estimates while not implicitly relying on treating the residuals as our best approximation to the "true" error structure? The answer is "yes." If we can make educated guesses about how the error structure for various development periods should look like (preferably in the shape of assumed parametric distributions), we can generate pseudo-data based on these educated guesses and then continue with calculating the resulting reserve and thus build up an empirical distribution of reserve estimates Monte Carlo style. This type of approach also avoids the practical limitations of Pearson or deviance residual bootstrapping we demonstrated in this paper.

5. CONCLUSIONS

Our case study of GLM-based bootstrapping for incremental development triangles reveals a serious limitation of the standard approach based on Pearson residuals: the possible existence of negative resampling values due to the linear nature of the rescaling procedure. We demonstrated that the obstacle of negative resampling value can be overcome by using deviance residuals that rely on non-liner rescaling. We also proved that deviance residuals based on the identity variance function cannot be used for resampling under all circumstances. A practical example of this was provided. In our discussion, we suggested that in a stochastic reserving context the main advantage of bootstrapping is that it can generate a distribution of reserve estimates that accounts for parameter correlations imposed by the estimation process. To use this advantage, we can also employ resampling schemes that are based on assumed parametric distributions and thus circumvent the limitations of non-parametric resampling revealed by this case study. Further research may also reveal more robust non-parametric resampling schemes.

Acknowledgment

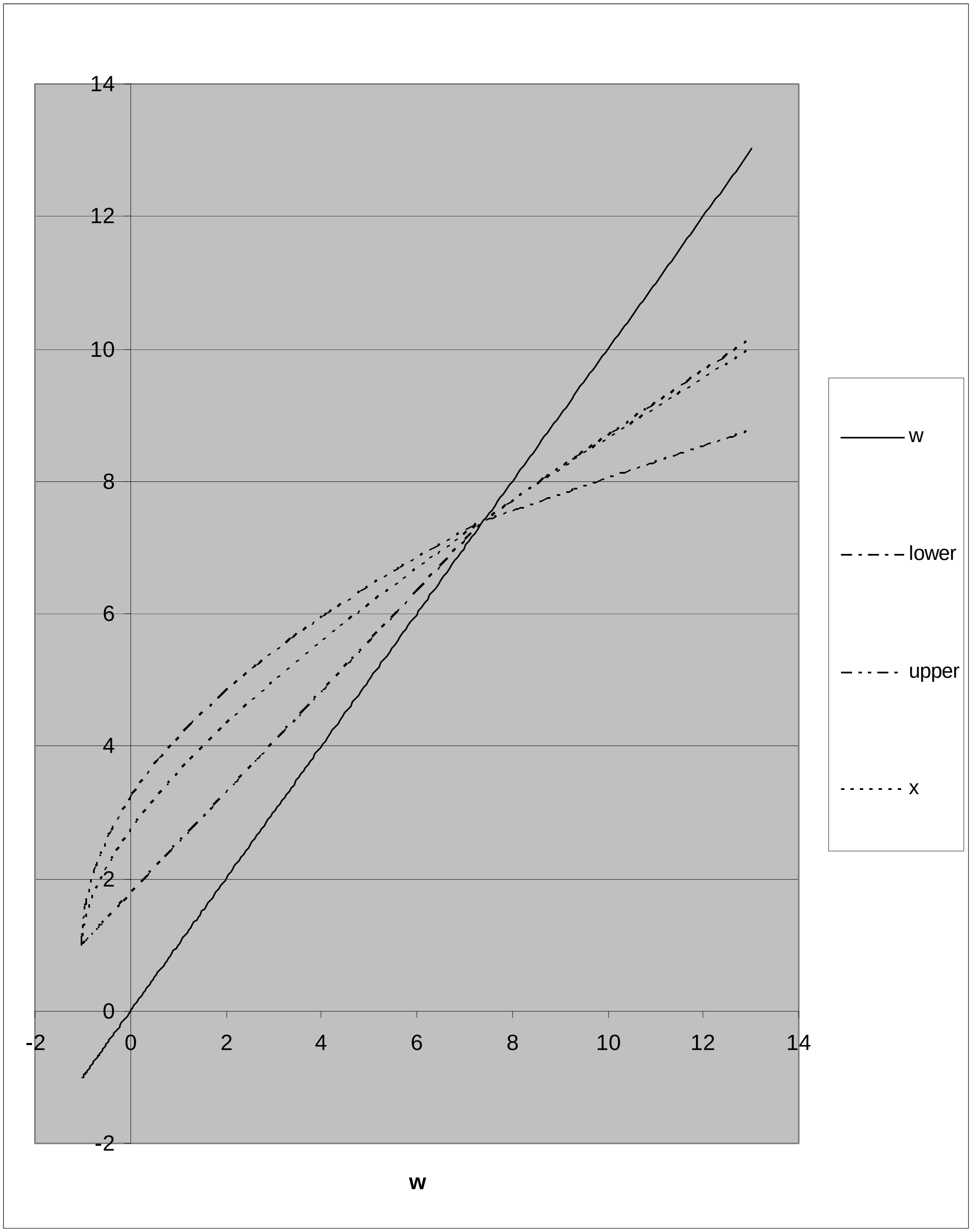
The companion MS Excel application to this paper is derived from a stochastic reserving model internally developed by PricewaterhouseCoopers' Actuarial and Insurance Management Solutions (AIMS) practice. The author thanks all PwC staff members who helped him with this development project. PwC AIMS also supported the author of this paper by providing him with writing time. Any errors, inaccuracies or views expressed in the paper are the author's responsibility alone. PwC's support for this project in no way represents an endorsement by PwC or PwC AIMS of any of the views expressed by the author or the methods presented in the paper.

APPENDIX A

In the case of positive deviance residuals we use the following upper and lower bounds to initialize the inversion algorithm:

w	Lower Bound	Upper Bound
-1	1	1
$\left(-1, e^2\right)$	$1+\frac{(w+1)\cdot\left(e^2-1\right)}{\left(e^2+1\right)}$	$1+\frac{\sqrt{w+1}\cdot\left(e^2-1\right)}{\sqrt{e^2+1}}$
e^2	e^2	e^2
$> e^2$	$\frac{w}{\ln(w)}+\frac{e^2}{2}$	$\frac{2\cdot w}{\ln(w)}$

We do not provide a formal proof that these bounds are valid, but *x*, *w*, and the lower and upper bounds can easily be plotted to visually demonstrate that this is case (see next page).


14
12
10
8
6
4
2
0
-2
-2
0
2
4
6
8
10
12
14
w
w
lower
upper
x

6. REFERENCES

[1] Anderson, D., et al., "A Practitioner's Guide to Generalized Linear Models—A CAS Study Note," 3rd ed., **2007**, http://www.casact.org/library/studynotes/anderson9.pdf.

[2] Barnett, G., and B. Zehnwirth, "Best Estimates for Reserves," *Proceedings of the Casualty Actuarial Society*, **2000**, Vol. 87, 245-303.

[3] Davison, A.C, and D.V. Hinkley, *Bootstrap Methods and Their Application*, Cambridge University Press, **1997**.

[4] England, P.D., and R.J. Verrall, "Predictive Distributions of Outstanding Liabilities in General Insurance," *Annals of Actuarial Science,* **2006**, Vol. 1, No 2, 221-270.

[5] Hartl, T., "Fitting a GLM to Incomplete Development Triangles," Casualty Actuarial Society *E-Forum*, **Summer 2010**.

[6] McCullagh, P., and J.A. Nelder, *Generalized Linear Models,* 2nd ed. (London: Chapman & Hall/CRC, **1989)**.

[7] Pinheiro, Paulo J R, et al., "Bootstrap Methodology in Claim Reserving," *Journal of Risk and Insurance*, **2003**, Vol. 70, No. 4, 701-714.

[8] Press, W.H., et al., *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, **1992**.

[9] Schmid, Frank A.., "Robust Loss Development Using MCMC," **2010**, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1501706.

[10] Wüthrich, Mario V., and Michael Merz, *Stochastic Claims Reserving Methods in Insurance* (Chichester, England: Wiley, **2008)**.

Abbreviations and notations

GLM, generalized linear model
LSQ, least squares
MLE, maximum likelihood estimator (or estimation)
iid, independent identically distributed

Biography of the Author

Thomas Hartl is a Manager within PricewaterhouseCoopers' Actuarial Insurance Management Solutions (AIMS) practice. He provides consulting services to insurance companies, reinsurers, and regulators. His responsibilities include the design, validation, and implementation of simulation models supporting statistical analysis for ERM, litigation support, predictive modeling, and stochastic reserving. Thomas Hartl is an Associate of the Casualty Actuarial Society, a Member of the American Academy of Actuaries, and holds a Ph.D. in mathematics from the University of Glasgow, Scotland.

Contact: thomas.hartl@us.pwc.com; 617-530-7524