risks

MDPI

Article

On Comparison of Stochastic Reserving Methods with Bootstrapping

Liivika Tee *, Meelis Käärik and Rauno Viin

Institute of Mathematics and Statistics, Faculty of Science and Technology, University of Tartu, J. Liivi 2,
50409 Tartu, Estonia; meelis.kaarik@ut.ee (M.K.); raunoviin@gmail.com (R.V.)
* Correspondence: liivika.tee@ut.ee; Tel.: +372-581-37299

Academic Editor: Mogens Steffensen
Received: 30 September 2016; Accepted: 20 December 2016; Published: 4 January 2017

Abstract: We consider the well-known stochastic reserve estimation methods on the basis of generalized linear models, such as the (over-dispersed) Poisson model, the gamma model and the log-normal model. For the likely variability of the claims reserve, bootstrap method is considered. In the bootstrapping framework, we discuss the choice of residuals, namely the Pearson residuals, the deviance residuals and the Anscombe residuals. In addition, several possible residual adjustments are discussed and compared in a case study. We carry out a practical implementation and comparison of methods using real-life insurance data to estimate reserves and their prediction errors. We propose to consider proper scoring rules for model validation, and the assessments will be drawn from an extensive case study.

Keywords: chain-ladder; claim reserving; bootstrap; generalized linear model; model validation

1. Introduction

Every non-life insurance company is obligated to compensate its policy holders for claims that meet the terms of the policy. In order to meet and administer its contractual obligations to policyholders, the insurance company has to set up loss reserves. Since loss events with the number and amount of claims are random, it is important to calculate the claims reserve carefully, as underestimation would lead to solvency problems, and overestimation unnecessarily holds the excess capital instead of using it for other purposes. The claims estimation is one of the basic actuarial tasks in the insurance industry, because it gives the certainty to be solvent at any time moment in the future. There is a variety of methods for the actuary to choose amongst for reserving purposes. The focus has mainly been on aggregate reserving techniques, where models perform analysis with aggregate claims data. In recent years, considerable attention has been given to stochastic micro-level models, which use claims-related data on an individual basis, rather than aggregating by underwriting year and development period (for a reference, see [1–3]). Despite the fact that stochastic micro-level models have emerged in an increasing steam of academic literature, these models are not substantially used by practitioners.

The most widely-used models are non-stochastic macro-level models, which are merely deterministic algorithms using aggregate claims data. The basic chain-ladder model is the flagship of macro-level models (see for details [4,5]). The simplest assumption of chain-ladder method is that payments will emerge in a similar way in each accident year. The proportionate increases in the known cumulative payments from one development year to the next can then be used to calculate the expected cumulative payments for future development years. Despite its well-known limitations, the chain-ladder remains as the most widely-applied claim reserving method, and several extensions of the model have been developed, for example the double chain-ladder method ([6]), which simultaneously uses a triangle of paid losses and a triangle of incurred claim counts, and the continuous chain-ladder method ([7]), which reformulates the triangular data as a histogram and proposes a continuous chain-ladder model through the use of a kernel smoother.

The chain-ladder method gives a point estimate, but the interest arises in developing estimates of the likely variability of the claims reserve. Stochastic macro-level models were first introduced in order to answer this question. An overview of stochastic macro-level models is given by [8,9]. A more thorough and detailed review is provided by [10]. Stochastic claims reserving starts with constructing a model that produces the actuary's best estimate and then using this model for estimating the prediction error of the model. Moreover, there is a tendency to find a model under which the best estimate is the one given by the chain-ladder. Within this group of models, the (over-dispersed) Poisson (ODP) model ([11]), gamma model ([12]), negative binomial model ([13]), log-normal model ([14]) and Mack's model ([15]) have received considerable attention. The first four models specify the distribution of the incremental losses, while the last is a distribution-free model that only specifies the first two moments.

The mean square error of prediction (MSEP), also known as the prediction error, has been used as the precision measure for the reserve estimates in most literature. It can be decomposed into two components: parameter uncertainty and process uncertainty (see, e.g., [8,16]). The former comes from the uncertainty in the estimation of parameters of the reserving model due to the limited sample size, whereas the latter comes from the intrinsic randomness of the claims development in the future. However, obtaining estimates for the standard error of prediction can be a difficult task. There are several analytical results for computing the prediction error (see [17]), but those estimates can be difficult to calculate or are only approximate values. For that reason, the advantage of the bootstrap technique can be taken. In addition, calculating the MSEP certainly provides great insight into the performance of reserve estimates, but other information such as the cash flow or risk measures are also of interest. Thus, for the full predictive distribution of reserve estimates, bootstrapping can be used for the solution. The bootstrap technique has been extensively studied in the claims reserving framework by various authors, such as [17–19]. The chain-ladder is still the benchmark for evaluating new models in the majority of the reserving literature. However, there is a need for more proper tools to validate and assess the quality of predictions when comparing different reserving methods. In order to validate the reserving method and identify any needed modifications, we need to rank the competing predictive models. We propose to consider scoring rules to measure the accuracy of probabilistic predictions.

The main purpose of this paper is to discuss different reserving methods on the basis of the chain-ladder method in combination with the bootstrap method in a case study approach. The definition of the proper residuals to base the bootstrap technique on is definitely an open subject when bootstrapping. We extend the work of [19] by using another useful type of residual with bootstrapping, and we carry out a comparative study among several stochastic models like the (over-dispersed) Poisson, the gamma and the log-normal model. We will use claims data from an Estonian insurance company for the case study, where we discuss the impact of the chosen models and the residuals on the reserve estimates and prediction errors. To evaluate the goodness of fit of the models, we carry out a model assessment.

The paper is set out as follows. In Section 2, we present a brief review of generalized linear models and their application to claim reserving, while in Section 3, we discuss some aspects linked to the bootstrap methodology. Section 4 is devoted to the application of the different methods to the real-life dataset. In Section 5, the comparative analysis for model validation with the Schedule P database is carried out. It is followed by the discussion in Section 6.

2. Chain-Ladder Method as a Generalized Linear Model

In this section, we introduce briefly the basic chain-ladder method, recall how the chain-ladder method is reformulated in the context of generalized linear models (GLM) and give a brief review of stochastic macro-level models, which will be used in the analysis. For a general introduction to GLM, we refer to [20].

Stochastic macro-level models use aggregate claims data, and some of the main advantages over non-stochastic macro-level models are the possibilities to obtain first two moments or the predictive distribution of the reserve estimate. Several often-used and traditional actuarial methods to complete a run-off triangle can be described by GLM. The actuarial literature has also shown a close connection between the chain-ladder method and the multiplicative Poisson model.

Without loss of generality, we assume that the data that have been collected for $i = 1, ..., n$ and $j = 1, ..., n$ consist of a triangle of incremental claims:

$$\left\{C_{ij} : \; i = 1, ..., n; \; j = 1, ..., n - i + 1\right\},$$

where the row index i refers to the year of origin and, depending on a particular situation, indicates the accident year, reporting year or underwriting year. The column index j refers to the development year, indicating the delay, more precisely loss disbursal, reporting year or accident year. Claims data are given as a run-off triangle as shown in Table 1.

Table 1. Run-off triangle with incremental claim amounts.

	Development Period *j*				
Year of Origin *i*	**1**	**2**	**3**	**...**	***n***
1	C_{11}	C_{12}	C_{13}	...	C_{1n}
2	C_{21}	C_{22}	...		
3	C_{31}	...			
⋮	⋮				
n	C_{n1}				

The cumulative claim amounts with accident year index i reported up to, and including, the delay index j are defined as:

$$D_{ij} = \sum_{k=1}^{j} C_{ik}.$$

Thus, D_{ij} is the total claims amount of accident year i, $i = 1, \ldots, n$, either paid or incurred up to development year j, $j = 1, \ldots, n$. The development factors of the chain-ladder technique are estimated as:

$$\widehat{\lambda}_j = \frac{\sum_{i=1}^{n-j} D_{i,j+1}}{\sum_{i=1}^{n-j} D_{ij}}, \quad j = 1, \ldots, n-1.$$

Generalized linear modeling is a methodology for modeling the relationships between variables. It generalizes the classical normal linear model, by relaxing some of its restrictive assumptions, and provides methods for the analysis of non-normal data. GLM is important in the analysis of insurance data, because with insurance data, the assumptions of the normal model are often not applicable. See [21] for a detailed description of generalized linear models for insurance data.

Following [11,19], the structure of the stochastic models for claim reserving in the terminology of GLM can be given by:

(1) incremental claim amounts C_{ij} belong to the exponential family,
(2) $E(C_{ij}) = \mu_{ij}$,
(3) $\eta_{ij} = g(\mu_{ij})$, where $g(\cdot)$ is the link function,
(4) linear predictor $\eta_{ij} = c + \tilde{\alpha}_i + \tilde{\beta}_j$ with an intercept c and factor effects $\tilde{\alpha}_i$ and $\tilde{\beta}_j$.

The given structure of GLM can be used to describe several often used actuarial methods. We consider the following multiplicative model ([9]), with a parameter for each row i, each column j and each diagonal $k = i + j - 1$:

$$C_{ij} \approx \alpha_i \cdot \beta_j \cdot \gamma_k, \tag{1}$$

where parameter α_i describes the effect of year of origin i, parameter β_j corresponds to development year j and γ_k describes the effect of calendar year $k = i + j - 1$. The approximation sign in Equation (1) expresses a difference caused by a chance, i.e., there is a possible deviation of the observation on the left-hand side from its mean value on the right-hand side. The model involves three time scales, which give rise to the well-known identification problem. Parametrization using three time scales has been introduced for instance by [22]. The identification problem has been revisited by several authors; see, for example, [23,24], who have proposed a canonical parametrization that is uniquely identified. In the framework of three time scales, we also face a problem with extrapolating the calendar estimates. Namely, we have no data on the values of γ_k for the future calendar years, e.g., if $k > n$. This can be overcome by assuming that the γ_k have a geometric pattern, with $\gamma_k \propto \gamma^k$ for some real number γ. Typically, the model (1) is simplified by taking $\gamma_k \equiv 1$, and the condition $\sum_{j=1}^{n} \beta_j = 1$ is imposed. If the parameters $\alpha_i > 0$ and β_j are estimated by using the maximum likelihood method, then the simplified model is a multiplicative GLM with log-link.

In the terminology of GLM, to linearize the multiplicative model (1), the logarithm is chosen as a link function (log-link). Hence:

$$\begin{aligned} E(C_{ij}) = \mu_{ij} &= \alpha_i \cdot \beta_j \cdot \gamma_k \\ &= \exp(\ln \alpha_i + \ln \beta_j + \ln \gamma_k), \end{aligned}$$

or, equivalently,

$$\ln E(C_{ij}) = \ln \alpha_i + \ln \beta_j + \ln \gamma_k. \tag{2}$$

Parameters of the given model are estimated by using the maximum likelihood method. After obtaining the estimates of the parameters, it is easy to complete the run-off triangle, simply by taking:

$$\widehat{C}_{ij} := \widehat{\alpha}_i \cdot \widehat{\beta}_j \cdot \widehat{\gamma}_k. \tag{3}$$

This simple model allows one to generate quite a few reserving techniques, depending on the assumptions set on the distribution of C_{ij}. It is common in claim reserving to consider the Poisson, gamma or log-normal distribution for the variable C_{ij}. We proceed with reviewing the following methods from Model (1).

The (over-dispersed) Poisson model: Already in 1975, a stochastic model corresponding to the Poisson model, which leads to the chain-ladder technique, was proposed. This model works on the incremental amounts C_{ij} from a Poisson distribution, where $E(C_{ij}) = \alpha_i \beta_j$ with unknown parameters α_i and β_j. Here, α_i is the expected ultimate claims amount (up to the latest development year so far observed), and β_j is the proportion of ultimate claims to emerge in each development year with the restriction $\sum_{k=1}^{n} \beta_k = 1$. The restriction immediately follows from the fact that β_j is interpreted as the proportion of claims reported in development year j. Obviously, the aggregate proportion over all periods has to be one.

We estimate the unknown parameters α_i and β_j from the triangle of known data with the maximum likelihood method. In the following, we use the notation Δ for the triangle of known data, i.e., the set of all (i, j), where C_{ij} is known. We also distinguish $\Delta_i = \{j : (i, j) \in \Delta\}$ and $\Delta^j = \{i : (i, j) \in \Delta\}$. The estimation procedure and results are given in the following lemma. The initial idea of the lemma is attributed to [12].

Lemma 1. *Assume that all C_{ij} are independent with a Poisson distribution, and $E(C_{ij}) = \alpha_i\beta_j$ holds. Then, the maximum likelihood estimators $\widehat{\alpha}_i$ and $\widehat{\beta}_j$ are given by:*

$$\widehat{\alpha}_i = \frac{\sum_{j\in\Delta_i} C_{ij}}{\sum_{j\in\Delta_i} \beta_j}, \quad i = 1, \ldots, n \tag{4}$$

and:

$$\widehat{\beta}_j = \frac{\sum_{i\in\Delta^j} C_{ij}}{\sum_{i\in\Delta^j} \alpha_i}, \quad j = 1, \ldots, n. \tag{5}$$

Proof of Lemma 1. We derive the maximum likelihood estimates for the unknown parameters α_i and β_j with the likelihood function:

$$L = \prod_{i,j\in\Delta} \frac{(\alpha_i\beta_j)^{C_{ij}}}{C_{ij}!} \exp(-\alpha_i\beta_j).$$

Therefore, the log likelihood function is:

$$\ell = \ln(L) = -\sum_{i,j\in\Delta} \alpha_i\beta_j + \sum_{i,j\in\Delta} C_{ij}\ln(\alpha_i\beta_j) - \sum_{i,j\in\Delta} \ln(C_{ij}!),$$

where the summation is for all i, j where C_{ij} is known. The maximum likelihood estimator consists of values of α_i, β_j, which maximize L or equivalently $\ln(L)$. They are given by the equations:

$$0 = \frac{\partial\ell}{\partial\alpha_i} = -\sum_{j\in\Delta_i} \beta_j + \sum_{j\in\Delta_i} C_{ij}\frac{1}{\alpha_i}, \quad i = 1, \ldots, n$$

and:

$$0 = \frac{\partial\ell}{\partial\beta_j} = -\sum_{i\in\Delta^j} \alpha_i + \sum_{i\in\Delta^j} C_{ij}\frac{1}{\beta_j}, \quad j = 1, \ldots, n.$$

Thus, the likelihood estimator α_i and β_j is given, respectively, by Formulas (4) and (5), and the lemma is proven. □

Thus, the proportion factors β_j express the ratio of the sum of observed incremental values for certain development year j with respect to certain ultimate claims, i.e., β_i denotes the proportion of claims reported in development year j. The parameters α_i refer to the ratio of the sum of observed incremental values for a certain origin year i to corresponding proportion factors. In other words, if the incremental claim amounts and respective proportions factors are known, it is simple to derive the corresponding ultimate claim α_i for origin year i. One can note the principal similarities with the chain-ladder technique, where development factors are also the outcomes of certain ratios.

The Poisson model can be cast into the form of a GLM, and to linearize the multiplicative model, we need to choose the logarithm as a link function, $\eta_{ij} = \ln(\mu_{ij})$, so that:

$$E(C_{ij}) = \mu_{ij} = \exp(\ln(\alpha_i) + \ln(\beta_j))$$

or, equivalently,

$$\ln(E(C_{ij})) = \ln(\alpha_i) + \ln(\beta_j) \tag{6}$$

where the structure of linear predictor (6) is still a chain-ladder type, because parameters for each row i and each column j are given. Hence, the structure (6) is defined as a GLM in which the incremental values C_{ij} are modeled as Poisson random variables with a log-link. Reparametrizing (6) gives us a structure of Property (4) defined in a GLM setting, i.e., we obtain a linear predictor:

$$\eta_{ij} = c + \tilde{\alpha}_i + \tilde{\beta}_j, \tag{7}$$

where parameter c can be considered as an intercept, which corresponds to the incremental amount in the cell (1, 1). This is obtained by taking:

$$\tilde{\alpha}_1 = \tilde{\beta}_1 = 0$$

to avoid over-parametrization. The Poisson model was studied in further detail by [25], where also a new canonical parametrization was proposed.

We recall that the only distributional assumptions used in GLMs are the functional mean-variance relationship and the fact that the distribution belongs to the exponential family. When defining a GLM, we can omit the distribution of C_{ij}'s and use only the most elementary information about the response variable, namely the relationship between variance and mean. This introduces a quasi-likelihood as an alternative, and using this elementary information alone can be often sufficient to stay close to the full efficiency of maximum likelihood estimators. Therefore, we can estimate the parameters by the maximum quasi-likelihood ([20]) instead of the maximum likelihood, and the estimators remain consistent. However, it is necessary to impose the constraint that the sum of the incremental claims in every row and column has to be non-negative. This means that quasi-likelihood could not be used, for instance, when modeling incurred data with a large number of negative incremental claims in the later development periods.

In the case of the Poisson distribution, the mentioned relationship is $Var(C_{ij}) = E(C_{ij})$, and allowing for more or less dispersion in the data can be generalized to $Var(C_{ij}) = \phi E(C_{ij})$ without any change in form and solution of the likelihood equations. This kind of generalization allows for more dispersion in the data, and one speaks of an over-dispersed Poisson (ODP) model. It is shown ([26]) that every ODP model can be transformed into the Poisson model by dividing all incremental claims by a certain parameter. The general form for the ODP model can be given as follows:

$$E(C_{ij}) = \mu_{ij} = \alpha_i \beta_j, \tag{8}$$

$$Var(C_{ij}) = \phi \alpha_i \beta_j, \tag{9}$$

where:

$$\sum_{k=1}^{n} \beta_k = 1.$$

The over-dispersion is introduced through the parameter ϕ, which is unknown and estimated from the data. Considering a single incremental payment C_{ij} with the origin year i and claim payments in development year j (yet to be observed), we obtain the estimates of future payments from the parameter estimates by inserting them into Equation (6) and exponentiating, resulting as:

$$\widehat{C}_{ij} = \widehat{\alpha}_i \widehat{\beta}_j = \exp(\widehat{\eta}_{ij}). \tag{10}$$

Given Equation (10), the reserve estimates for any origin year can be derived by:

$$\widehat{R}_i = \widehat{\alpha}_i \widehat{\beta}_{n+2-i} + \ldots + \widehat{\alpha}_i \widehat{\beta}_n, \quad i = 2, \ldots, n, \tag{11}$$

and the reserve estimate for the total amount can be easily derived by summation:

$$\widehat{R} = \sum_{i=2}^{n} \widehat{R}_i = \sum_{i=2}^{n} (\widehat{\alpha}_i \widehat{\beta}_{n+2-i} + \ldots + \widehat{\alpha}_i \widehat{\beta}_n), \quad i = 2, \ldots, n. \tag{12}$$

The negative binomial model can be derived from the Poisson model, and thus, these models are very closely related, but with a different parameterization. The model was first derived by [13], by integrating out the row parameters from the Poisson model. The predictive distributions of both models are basically the same and give identical predicted values.

Log-normal model: When considering the log-normal distribution to describe claim amounts (see for a reference [14]), we can still continue to use GLM for the logs of the incremental claim amounts. The log-normal class of models are given as:

$$\ln(C_{ij}) \sim N(\mu_{ij}, \sigma^2),$$

i.e.,

$$E(\ln(C_{ij})) = \mu_{ij} \text{ and } Var(\ln(C_{ij})) = \sigma^2.$$

Now, the identity link function is used, and the normal responses $\ln(C_{ij})$ are assumed to decompose (additively) into a deterministic non-random component with mean $\mu_{ij} = \eta_{ij}$ and normally-distributed random error components with zero mean.

Following [8], the fitted values on a log scale, given the estimates for the parameters in the linear predictor η_{ij} and the process variance σ^2, are obtained by forming the appropriate sum of estimates. Obtaining the estimates for the mean on the untransformed scale is not that simple. We cannot just exponentiate the linear predictor, since that would give an estimate of the median. Therefore, the fitted values on the untransformed scale are given by:

$$\widehat{C}_{ij} = \exp(\widehat{\eta}_{ij} + \frac{1}{2}\widehat{\sigma}^2_{ij}), \tag{13}$$

which is in the standard form of the expected value of a log-normal distribution and where:

$$\widehat{\sigma}^2_{ij} = Var(\widehat{\eta}_{ij}) + \widehat{\sigma}^2$$

are the prediction variance of the linear predictor. With already familiar notation (from the ODP subsection), we denote the triangle of predicted claims contributing to the reserve estimates by ∇. The reserve estimate in origin year i is given by summing the predicted values in row i of ∇, i.e., $\widehat{R}_i = \sum_{j \in \nabla_i} \widehat{C}_{ij}$, and the total reserve estimate, summing the predicted values in row i and in column j of ∇, is given by $\widehat{R} = \sum_{i,j \in \nabla} \widehat{C}_{ij}$. The log-normal model is also referred to as the geometric chain-ladder model; see this additional analysis in [27].

Gamma model: A further model was proposed by [12] with a multiplicative parametric structure for the mean incremental claims amounts, which are modeled as gamma response variables. As noted in [11], the same model can be fitted using the GLM described in over-dispersed Poisson model, but in which the incremental claim amounts are modeled as independent gamma response variables with a logarithmic link function and the same linear predictor and require a change in (9). As with the log-normal model, the predicted values provided by the gamma model are usually close to the chain-ladder estimates, but it cannot be guaranteed. The gamma model implemented as a generalized linear model gives exactly the same reserve estimates as the gamma model implemented by [12]. The gamma model is given with the mean:

$$E(C_{ij}) = \mu_{ij} = \alpha_i \beta_j,$$

and with the variance:

$$Var(C_{ij}) = \phi(E(C_{ij}))^2 = \phi\mu^2_{ij}.$$

To obtain reserve estimates with the gamma model for any origin year or for the overall amount, the same formulas as defined in the ODP model, (11) and (12), respectively, can be used. The limitation of both the gamma and ODP model is that each incremental value should be nonnegative.

3. The Bootstrap Technique

Bootstrapping is a popular technique in stochastic claims reserving because of the simplicity and flexibility of the approach. We are using bootstrapping to estimate the prediction error and to approximate the predictive distribution. An analytical derivation of the prediction error of the total reserve estimate may be preferable from a theoretical perspective, but it is often impracticable due to complex reserve estimators. For the classical chain-ladder method, [15] derived an analytical expression of the MSEP within an autoregressive formulation of the claims development using a second-moment assumption. The first order Taylor approximation of the corresponding MSEP within the GLM framework was derived by [17]. As said, known theoretical estimators are difficult to calculate and are still merely approximate values.

For both the classical and generalized linear model, it is common to adopt either a paired bootstrap where resampling is done directly from the observations or the residuals bootstrap where resampling is applied to the residuals of the model. The paired bootstrap is more robust than the residual bootstrap, but only the residual bootstrap can be implemented in the context of the claim reserving, given the dependence between some observations and the parameter estimates. If the type of residuals adopted is the same, then mixing GLMs with bootstrapping is similar to combining the chain-ladder method with bootstrapping. The residuals obtained from applying a GLM to the past claims data are used in the resampling process of bootstrapping. With each re-sampled set of residuals, an upper triangle can be constructed, and the stochastic chain-ladder can be applied again. The lower triangle is then simulated from the assumed distribution with the first two moments determined by the stochastic chain-ladder. Thereafter, an empirical distribution is formed, from which the required inferences can be drawn.

3.1. Residuals

The process of creating a distribution for the reserve can be done by bootstrapping, either parametric bootstrapping or non-parametric. It is common to use the residuals to bootstrap a claims reserves distribution, which is a non-parametric bootstrapping method. For GLMs, the main reason for not simply examining the raw residuals is the difficulty of checking the validity of the assumed mean-variance relationship from the raw residuals.

One of the most used residuals in model diagnostics are the Pearson residuals and the deviance residuals. Furthermore another residual, the Anscombe residual, is often mentioned as a possible residual to consider, but is rarely applied in further work due to being known as a less commonly-used residual. However, following [21], the Anscombe and the deviance residuals are mathematically different, but numerically, they give similar results. The Anscombe residual tries to make the residuals "as close to normal as possible", and given that the response distribution has been correctly specified, the deviance residuals are also approximately normally distributed. Thus, contrary to the usual practice, we explore in the following the use of the Anscombe residuals. A version of the deletion residual is also available under the GLM setting, which is related to the Pearson residual, but their forms are rather complicated and, thus, omitted.

The Pearson residuals are just rescaled versions of the raw or response residuals and are defined as:

$$r_{ij}^{P} = \frac{C_{ij} - \widehat{\mu}_{ij}}{\sqrt{V(\widehat{\mu}_{ij})}},$$

where $V(\cdot)$ is a variance function. The Pearson residuals need to be adjusted in order to obtain (approximately) equal variance, and there are different adjustments suggested by several authors. It was proposed by [8,17] to adjust the residuals by multiplying them by a correction factor:

$$r_{ij}^{PE} = \sqrt{\frac{n}{n-p}} r_{ij}^{P},$$

where n is the sample size and p is the number of estimated parameters. In correspondence with the classical linear model, often the "hat" matrix of the model is used to standardize the Pearson residuals, which are given as:

$$r_{ij}^{P*} = \frac{r_{ij}^{P}}{\sqrt{\widehat{\phi}(1-h_{ij})}}, \tag{14}$$

where $\widehat{\phi}$ is a scale parameter estimated from the data, and the factor h_{ij} is the corresponding element of the diagonal of the "hat" matrix. This matrix is given for classical linear models by $\mathbf{H} = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}$, and it can be generalized for GLM as follows:

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W},$$

where $\mathbf{X}$ is a design matrix and $\mathbf{W}$ is a diagonal matrix with elements:

$$w_{ii} = \left(V(\mu_{ij}) \left(\frac{\partial \eta_{ij}}{\partial \mu_{ij}} \right)^{2} \right)^{-1}$$

on the diagonal (see [20] for details).

The distribution of Pearson residuals for non-normal distributions is often markedly skewed and, thus, may fail to have properties similar to those of a normal-theory residual. Then, the Anscombe residual can be a good alternative to the Pearson residual. The Anscombe residuals do not use the variable C_{ij} directly, but instead a transformation $A(C_{ij})$. The function $A(\cdot)$ is chosen to make the distribution of $A(C_{ij})$ as normal as possible and in the context of GLM the Anscombe residual is defined as

$$r_{ij}^{A} = \frac{A(C_{ij}) - A(\widehat{\mu}_{ij})}{A'(\widehat{\mu}_{ij}) \cdot \sqrt{V(\widehat{\mu}_{ij})}},$$

where $A'(\mu)$ is the derivative of $A(\mu)$ and $V(t)$ is the variance function. For the Poisson model, the Anscombe residuals are defined by:

$$r_{ij}^{A} = \frac{\frac{3}{2}\left(C_{ij}^{\frac{2}{3}} - \widehat{\mu}_{ij}^{\frac{2}{3}} \right)}{\widehat{\mu}_{ij}^{\frac{1}{6}}}$$

and for the gamma model the residuals are defined as:

$$r_{ij}^{A} = 3\left(\left(\frac{C_{ij}}{\widehat{\mu}_{ij}} \right)^{\frac{1}{3}} - 1 \right).$$

It is easy to see that in case of the normal model, the Anscombe residuals are equivalent to the classical residuals, and thus, for the log-normal model the residuals are defined as:

$$r_{ij}^{A} = \ln(C_{ij}) - \widehat{\mu}_{ij}$$

since $V(\mu_{ij}) = 1$. For a detailed overview of residuals, see [21].

Prediction errors with the bootstrapping method are compared based on the type of residuals used and if or how we have adjusted the residuals. It is important to notice that the residuals of the calculated values of the first column in the last row and of the first row in the last column are always equal to zero, i.e., $\widehat{\mu}_{1n} - C_{1n} = 0$ and $\widehat{\mu}_{n1} - C_{n1} = 0$. These are zeros due to the defined linear structure adopted in the models implying the estimates for some of the parameters depend on one observation only. The reason for the correction of zeros is that the bootstrap method assumes the random variables (in this case, residuals) to be i.i.d. random variables, but in this case, there are two non-random residuals, which are always fixed as zeros. Thus, we remove zero-residuals and replace them with residuals resampled from the remaining ones. In this paper, we consider the Pearson and Anscombe residuals first without corrections, then with the zeros corrected and lastly standardized versions of residuals.

3.2. Prediction Error and Confidence Limits

A commonly-used measure of variability is the prediction error. In this context, we use the expected value as the prediction. The prediction error consists of two parts: the process variance and the estimation variance. The mean squared error of the prediction (MSEP) $\widehat{C}_{ij}$ is given by:

$$\begin{aligned} MSEP(\widehat{C}_{ij}) &= E((C_{ij} - \widehat{C}_{ij})^2) \\ &\approx (E(C_{ij}) - E(\widehat{C}_{ij}))^2 + Var(C_{ij} - \widehat{C}_{ij}) \\ &= Var(C_{ij}) + Var(\widehat{C}_{ij}), \end{aligned} \tag{15}$$

where $Var(C_{ij})$ denotes the process variance and $Var(\widehat{C}_{ij})$ denotes the estimation variance. Equation (15) is valid for the over-dispersed Poisson, the gamma and the log-normal reserving models. Both terms have explicit expressions depending on which prediction model is used; see for instance [17,28]. The reserve estimate for origin year i is given by the sum of the predicted values in row i of Δ, i.e., $\widehat{R}_i = \sum_{j \in \Delta_i} \widehat{C}_{ij}$, and for the estimate of the total, reserve Formula (12) can be used. The calculation of prediction errors for origin year reserve estimates and overall reserve estimates require more effort. Predicted values in each row are based on the same parameters, and predicted values in the same column are based on the same parameters; thus, we need to handle dependency. The variance of the sum of predicted values is considered, taking into account any covariances between predicted values. Under certain assumptions, we need to consider only covariances arising in the estimation variance. For detailed derivations of prediction errors for different models, we refer to [17]. All of these components can be rather difficult to calculate analytically, whereas the bootstrap procedure is practically prudent and does not require the summation of a large collection of terms, unlike the analytic and distribution-free approaches.

One possible bootstrap prediction approach takes the advantage of the central limit theorem by approximating the distribution of the reserve by means of a normal distribution with the expected value given by the initial forecast (with the original data) and the standard deviation given by the standard error of prediction, which is an estimate of the square root of the estimation variance. However, it cannot be compared directly with the analytic equivalent since the bootstrap standard error does not take into account the number of parameters used in fitting the model, i.e., the bootstrap process simply uses the residuals with no regard as to how they are obtained. As suggested by [17], the appropriate adjustment to the bootstrap estimation variance to take account of the number of parameters estimated is to multiply the bootstrap estimation variance by $\frac{n}{n-p}$.

The analytic estimates of the estimation variance involve variance and covariance terms, which implicitly include the scale parameter ϕ in their calculation. The scale parameter can be estimated, for example, as the Pearson chi-squared statistic divided by the degrees of freedom:

$$\phi_P = \frac{\sum (r_{ij}^P)^2}{n-p},$$

where n is the number of data points in the sample and p is the number of parameters estimated, and the summation is over the number of residuals. The bootstrap prediction error is the square root of the sum of the squares of estimation variance and process variance,

$$SEP_{bs}(R) = \sqrt{Var(C_{ij}) + \frac{n}{n-p}(SE_{bs}(R))^2},$$

where R stands for the total reserve (but the formula can be applied analogously in case of origin year reserves). $SE_{bs}(R)$ is the bootstrap standard error of the reserve estimate, and process variance $Var(C_{ij})$ has an explicit form depending on the considered model. In the case of the ODP model and gamma model, the process variance would be:

$$\sum_{i,j\in\Delta} \phi_P \mu_{ij} = \phi_P \sum_{i,j\in\Delta} \mu_{ij} = \phi_P R,$$

and:

$$\sum_{i,j\in\Delta} \phi_P \mu_{ij}^2 = \phi_P \sum_{i,j\in\Delta} \mu_{ij}^2,$$

respectively. In the case of the log-normal model, the process variance is simply σ^2.

Following [19,29], we consider an alternative bootstrapping procedure to obtain an upper confidence limit for the forecasts of the aggregate values. This approach (in the following named the PPE-method) includes two resampling procedures in the same bootstrap "iteration", but the results should be more robust against deviations from the hypothesis of the model. The idea is to define an adequate prediction error as a function of the bootstrap estimate and a bootstrap simulation of the future reality and to record the value of this prediction error for each bootstrap "iteration". Then, use the desired percentile of this prediction error, and combine it with the initial prediction to obtain the upper limit of the prediction interval. See [19] for the step-by-step explanation of this alternative approach.

4. Case Study

To enable a comparison with previously-discussed methods in the framework of bootstrapping with defined residuals, we use the real-life dataset from an Estonian insurance company. The data considered describe the paid out claims and are shown here in incremental form. We are interested in the impact of the choice of the models and, mainly, in the effect of the choice of residuals and its adjustments.

We use both the Pearson and the Anscombe residuals first without corrections, then with the zeros corrected and lastly standardized residuals together with the zero correction. It is clear that using just standardized residuals will lead to the same results as obtained with the zero-corrected residuals; thus, we do not consider standardized residuals independently in the comparative study. In addition, we compare the obtained prediction errors and obtain the upper limits using both bootstrap approaches, i.e., the regular SEP-method based on the standard error of prediction and the alternative (using pseudo-reality) PPE method. We present PPE prediction errors only for the total reserve. When comparing SEP and PPE prediction errors, we have to take into account that different units are used: SEP prediction error equals one standard deviation, and PPE prediction error equals (approximately) 1.645 standard deviations (95%-quantile of normal distribution). This means that we have to multiply the prediction error obtained with SEP method by 1.645 and add it to the reserve estimate to obtain an upper confidence limit for the total reserve with the SEP method. In the case of the PPE method, we simply sum the prediction error and the mean to obtain the upper limit.

Reserve estimates provided by the over-dispersed Poisson model, the gamma model and log-normal model using the GLM implementation in the framework of bootstrapping with residuals outlined in this paper are shown in Tables 3–7 below. As one can see, the data considered are rather inconvenient (see Table 2), i.e., the large fluctuation of the values in the triangle is obvious: the smallest incremental value is 1022, and the largest one is 10,660,074, which is a 10,430-fold difference. The second column in Tables 3–7 shows a point estimate for the reserve. These estimates are obtained directly from the defined model (not depending on the bootstrap procedure), and the point estimates do not depend on the choice of residual or on its correction.

The most problematic stage in the bootstrap method is the formation of the pseudo-data. If the magnitudes of the incremental values differ significantly, it is quite likely that the values of simulated residuals (simulated from the initial set of residuals) are sufficiently high compared to the predicted incremental values to cause the negative values to appear in the (pseudo-)data due to the use of the inverse function. Most of the probability distributions used in loss reserving are non-negative (or positive) valued; thus, the problem with negative values in the (pseudo-)data can often appear. For example, in the case of the Poisson distribution, the negative incremental values are often replaced by zeros in practice. Since incremental values in Table 2 have a high volatility, we experienced some negative incremental values in the pseudo-data when using the gamma model with the Pearson residuals. We also tried to replace the appearing negative values with ones, but that caused non-convergence of the parameters. Thus, we could not present the results of the gamma model and the Pearson residuals with the given dataset. There were no problems in the case of the Anscombe residuals. See Table 8 for an overview of the experienced negative values in the pseudo-data for each considered model and residual adjustment with the given dataset in Table 2.

Table 2. Full run-off triangle for paid out claims.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
2000	4,734,994	1,885,305	281,240	504,341	524,449	365,049	100,761	32,449	3697	56,901
2001	4,344,093	1,783,774	243,849	339,985	49,482	178,961	508,272	78,125	1022	
2002	5,288,867	1,795,855	303,246	351,320	316,038	33,501	88,774	31,102		
2003	5,357,617	2,548,383	336,749	403,501	348,378	236,017	12,982			
2004	5,737,732	2,574,724	971,320	280,140	226,212	152,127				
2005	5,635,064	2,758,392	241,734	268,113	429,503					
2006	6,629,504	3,045,252	356,119	200,420						
2007	6,824,829	2,669,579	166,400							
2008	8,116,439	3,428,535								
2009	10,660,074									

We first have a look at the results obtained by ODP model with using the Pearson residuals (see Table 3) and the Anscombe residuals (Table 4). The tables present the point estimates along with the standard errors of prediction for the three situations considered, as well as the upper limits for a confidence level of 95%. The standard errors of prediction grow up if we introduce the zero corrections, and consequently, the same happens to the upper limits, but the same estimates drop if we use the standardization (see Formula (14)) with zero correction. The prediction errors (SEP) in the case of the Poisson model with Pearson residuals are varying from 1.6 million–1.94 million, depending on the residual adjustment, whereas in the case of the Anscombe residuals (see Table 4), the prediction errors vary from 1.47 million–1.76 million. This means that the 95% confidence limits for the total reserve prediction are between 16 million and 16.6 million in the case of the Pearson residuals and

15.8 million and 16.3 million in the case of the Anscombe residuals, given the Poisson model and residual adjustments.

Table 3. Over-dispersed Poisson model with Pearson residuals.

Year	Est. Reserve	Without Corrections		Zero-Correction		Zero-Correction & Stand.	
		SEP	**Upper 95%**	**SEP**	**Upper 95%**	**SEP**	**Upper 95%**
2	50,796	90,377	199,466	89,795	198,509	69,858	165,713
3	57,837	97,791	218,702	97,051	217,485	74,943	181,117
4	120,029	135,467	342,872	135,571	343,043	115,039	309,268
5	348,993	220,918	712,403	225,328	719,658	207,318	690,031
6	552,215	271,860	999,425	270,829	997,728	259,708	979,434
7	1,024,516	374,459	1,640,501	381,686	1,652,389	361,525	1,619,225
8	1,406,290	441,811	2,133,069	444,127	2,136,879	421,587	2,099,800
9	2,283,616	576,547	3,232,037	578,861	3,235,843	549,794	3,188,028
10	7,560,816	1,264,024	9,640,136	1,249,066	9,615,530	979,054	9,171,360
Total	**13,405,108**	**1,944,083**	**16,603,125**	**1,944,997**	**16,604,628**	**1,603,405**	**16,042,710**
PPE		3,182,150	16,587,258	3,082,305	16,487,413	1,625,348	15,030,456
PPE/SEP		1.637	0.999	1.585	0.993	1.014	0.937

Table 4. Over-dispersed Poisson model with Anscombe residuals.

Year	Est. Reserve	Without Corrections		Zero-Correction		Zero-Correction & Stand.	
		SEP	**Upper 95%**	**SEP**	**Upper 95%**	**SEP**	**Upper 95%**
2	50,796	85,484	191,418	87,509	194,748	69,751	165,536
3	57,837	91,878	208,975	94,369	213,073	75,139	181,440
4	120,029	129,886	333,690	133,400	339,471	114,032	307,612
5	348,993	211,456	696,839	216,857	705,723	200,673	679,101
6	552,215	260,652	980,988	262,344	983,771	250,894	964,937
7	1,024,516	357,010	1,611,798	364,678	1,624,412	347,598	1,596,314
8	1,406,290	415,877	2,090,406	421,084	2,098,972	406,742	2,075,380
9	2,283,616	542,459	3,175,961	544,458	3,179,249	522,033	3,142,361
10	7,560,816	1,124,403	9,410,459	1,109,368	9,385,726	934,009	9,097,261
Total	**13,405,108**	**1,727,161**	**16,246,288**	**1,758,340**	**16,297,578**	**1,469,680**	**15,822,732**
PPE		2,029,479	15,434,588	2,004,348	15,409,456	1,275,538	14,680,646
PPE/SEP		1.175	0.950	1.140	0.946	0.868	0.928

Using the Anscombe residuals, the same pattern of changes of prediction errors (and also upper limits) can be seen, but the prediction errors, as previously said, are smaller than the Pearson residuals. We see that the zero corrections do not effect the prediction errors significantly. The prediction errors without any corrections with the Anscombe residuals are 13% smaller than with the Pearson residuals. The corresponding numbers with zero correction and zero correction with standardization are 10% and 8%, respectively.

To compare the behavior of two bootstrapping approaches, we have the last two lines of each table presenting the prediction errors and the upper confidence limits of the total reserve obtained by the PPE method and the ratio of the results by the PPE method and SEP method. We can see that the upper confidence limits for the total reserve are lower with the PPE method (all of the ratios $\frac{PPE}{SEP}$ are smaller than one). On the other hand, the prediction errors (depending on the residual adjustments) obtained by the PPE method are higher than the estimates obtained by the SEP method. However, the ratios seem to decrease if we correct the residuals. In the case of Pearson residuals with zero correction and standardization, the corresponding ratio is slightly over one, and in the case of the Anscombe residuals, it is slightly below one.

Fitting the gamma model gives similar, but not identical, reserve estimates (see Table 5) compared to the results obtained by ODP. The point estimate for the total reserve with the gamma model is

12.1 million, whereas with the Poisson model it was 13.4, which is 10.7% higher. If we compare the reserve estimates by origin year, then the biggest difference can be seen on the third year, where the difference is 55.8%. In the case of the gamma model and the Anscombe residuals, the prediction errors for the total reserve vary from 3.35 million–5.04 million. The upper limit for the total reserve in the case of the gamma model reaches 20.43 million. In a nutshell, when comparing the Poisson and the gamma model in this particular dataset, the latter gives us a smaller total reserve estimate, but higher prediction errors and, thus, higher upper limits for the reserve. In the case of both models, the PPE method tends to give higher prediction errors, except the case when the residuals are zero-corrected and standardized.

Table 5. Gamma model with Anscombe residuals.

Year	Est. Reserve	Without Corrections		Zero-Correction		Zero-Correction & Stand.	
		SEP	**Upper 95%**	**SEP**	**Upper 95%**	**SEP**	**Upper 95%**
2	50,012	38,160	112,785	39,008	114,180	30,965	100,950
3	37,119	26,904	81,376	27,827	82,895	22,081	73,442
4	93,433	48,396	173,045	49,667	175,135	44,190	166,126
5	332,152	159,500	594,530	162,956	600,215	161,306	597,501
6	454,013	193,496	772,314	197,412	778,756	198,066	779,831
7	782,169	329,614	1,324,384	324,932	1,316,682	324,997	1,316,788
8	1,031,664	423,941	1,729,046	438,924	1,753,693	429,318	1,737,892
9	2,090,955	945,444	3,646,210	974,441	3,693,911	879,649	3,537,977
10	7,270,705	4,520,261	14,706,534	4,810,081	15,183,288	3,060,442	12,305,132
Total	**12,142,220**	**4,692,325**	**19,861,095**	**5,038,337**	**20,430,285**	**3,356,603**	**17,663,832**
PPE		7,586,523	19,728,743	6,993,945	19,136,166	2,839,397	14,981,617
PPE/SEP		1.617	0.993	1.388	0.937	0.846	0.848

From Tables 6 and 7, we can see the results of the log-normal model. The point estimate among all of the considered models is the lowest with the log-normal model, namely 10.8 million. However, we note a high increase in the prediction errors, especially in the case of residual's zero correction.

Table 6. Log-normal model with Pearson residuals.

Year	Est. Reserve	Without Corrections		Zero-Correction		Zero-Correction & Stand.	
		SEP	**Upper 95%**	**SEP**	**Upper 95%**	**SEP**	**Upper 95%**
2	42,904	52,003	128,449	51,823	128,152	15,151	67,827
3	36,824	39,224	101,347	47,523	114,999	13,957	59,783
4	80,170	57,605	174,930	63,622	184,828	31,949	132,726
5	215,413	107,603	391,661	120,549	413,716	94,383	370,673
6	351,163	166,083	624,369	172,026	634,146	162,592	618,626
7	600,400	290,000	1,077,450	288,431	1,074,868	281,236	1,063,033
8	819,029	422,285	1,513,687	431,693	1,529,163	406,269	1,487,341
9	1,790,227	1,254,931	3,854,588	1,437,501	4,154,916	1,092,042	3,586,636
10	6,871,745	8,625,252	21,060,284	10,534,588	24,201,142	2,188,477	10,471,789
Total	**10,807,874**	**8,751,120**	**25,203,481**	**10,741,298**	**28,477,309**	**2,696,996**	**15,244,432**
PPE		6,591,299	17,399,173	6,084,858	16,892,732	3,153,855	13,961,729
PPE/SEP		0.753	0.690	0.566	0.593	1.169	0.916

Table 7. Log-normal model with Anscombe residuals.

Year	Est. Reserve	Without Corrections		Zero-Correction		Zero-Correction & Stand.	
		SEP	Upper 95%	SEP	Upper 95%	SEP	Upper 95%
2	42,904	34,201	99,164	34,575	99,779	11,889	62,461
3	36,824	26,572	80,534	31,038	87,881	10,885	54,729
4	80,170	39,629	145,359	43,036	150,964	24,827	121,010
5	215,413	79,247	345,774	86,722	358,152	72,523	334,713
6	351,163	124,969	556,737	128,456	562,473	124,209	555,486
7	600,400	221,705	965,104	220,991	963,930	217,118	957,559
8	819,029	321,006	1,347,083	328,293	1,359,070	311,412	1,331,301
9	1,790,227	924,401	3,310,866	1,014,638	3,459,306	804,123	3,113,009
10	6,871,745	5,677,582	16,211,367	6,631,999	17,781,383	1,712,612	9,688,991
Total	**10,807,874**	**5,782,386**	**20,319,898**	**6,793,931**	**21,983,890**	**2,082,248**	**14,233,171**
PPE		5,934,922	16,742,796	5,279,051	16,086,925	2,693,775	13,501,649
PPE/SEP		1.026	0.824	0.0.777	0.731	1.294	0.949

The prediction errors for the total reserve with the log-normal model with the Pearson residuals vary from 2.7 million–10.7 million, depending on the residual's adjustments. The upper limits for the total reserve with the Pearson residuals vary from 15.2 million–28.47 million; this shows a great fluctuation of the estimates. The prediction errors with the Anscombe residuals are between two million and 6.8 million; thus, the 95% confidence limit for the total reserve is between 14.2 million and 22 million, depending on the residual's adjustments. However, higher values of the prediction errors should not be surprising, as the log-normal model is a more "conservative" model than, for example, the Poisson model or the gamma model. The prediction errors as the % of the total reserve estimates obtained by the Pearson residuals without corrections, with zeros corrected and then with zero correction with standardization are 81%, 99% and 25%, respectively. The corresponding % of prediction errors in the case of the Anscombe residuals are 53%, 63% and 19%, respectively. We see that the same pattern follows as before; if we use zero correction, then the prediction errors (and consequently, the upper limits, as well) are the highest. The lowest prediction errors are obtained by the zero correction together with using standardization. Furthermore, in case of the log-normal model, we see that the PPE method gives smaller upper limits than the SEP method for the total reserve. Note that when it comes to the prediction errors, the PPE method does not continue to give higher prediction estimates than the SEP method, which was the case with the Poisson and the gamma models. We see from the Tables 3–7 that on the 10th year, the estimated reserve is the highest and is approximately three-times higher than the estimated reserve on the previous year. The reserve estimate on the 10th year makes nearly 56.4% of the total reserve estimate in the case of the Poisson model, 59.9% in the case of the gamma model and 63.4% in the case of the log-normal model, which is the highest percentage. This high proportion of the reserve estimate on one particular year can be explained by having a look at the initial dataset, Table 2, where we see that on the last year, 2009, we have the largest value in the whole dataset.

We can draw four main conclusions from analyzing this dataset:

1. The over-dispersed Poisson model produces the highest estimated claim reserve, and the log-normal model produces the smallest estimated claim reserves. The figures of the gamma model are not that different from the ODP model.
2. The standard errors of prediction are quite different and consequently the estimated upper limits. These differences tend to be greater especially on the first years, since estimations are based on few predictions. The highest prediction errors are produced by the log-normal model, and the lowest prediction errors were obtained by the over-dispersed Poisson model.

3. With this particular dataset, the prediction errors are the lowest with the Anscombe residuals. Furthermore, no matter which residual of the two is used, the lowest prediction errors are obtained by using the zero correction with standardization.
4. When comparing the two bootstrap procedures, we can conclude that using the (alternative) PPE method, the upper confidence limits for the total reserve are lower with each considered model.

As we mentioned beforehand the possible problem associated with the negative values in the pseudo-data, we present Table 8, which gives an overview of the amount of the negative values appearing in the procedure of creating a pseudo-data in the case of 1000 iterations. Roughly speaking, we observe that with the Poisson Models 1–2, negative pseudo-incremental values appeared with every iteration step. This is rather expected since the incremental values in the data differ largely. Note that using the Pearson residual caused more negative values than using the Anscombe residuals. There were no negative values in the pseudo-data in the case of the gamma, nor the log-normal model.

Table 8. Amount of negative values appearing in the pseudo-data during 1000 iterations.

Type of Adjustment	Poisson Model		Log-Normal Model		Gamma Model
	Pearson	Anscombe	Pearson	Anscombe	Anscombe
Without corrections	2281	1132	0	0	0
Zero-correction	2314	1172	0	0	0
Zero-correction & Stand.	2207	912	0	0	0

The presented prediction errors in the Tables 3–7 above helped us to compare the variability of the mean of the total reserve. However, it can be also helpful to have an idea of the upper limit of the total reserve in general. The quantiles for a random total reserve and for the mean of the total reserve in the case of the Poisson model are presented in the tables below, Tables 9 and 10.

Table 9. The upper confidence limits for the total reserve and for the mean of the total reserve by the Poisson model and Pearson residuals.

The Quantile		The Adjustment Type		
		Without Corrections	Zero-Correction	Zero-Correction & Stand.
90%	Reserve	16,189,247	16,284,477	15,903,554
	Mean	15,893,535	15,894,705	15,457,467
95%	Reserve	17,046,324	17,712,940	17,141,555
	Mean	16,603,125	16,604,628	16,042,710
99%	Reserve	18,476,692	18,570,018	17,712,940
	Mean	17,934,822	17,936,951	17,141,042

As expected, the upper limits of the mean of the total reserve are lower than the upper limit of the random total payment (reserve). The adjustments of the residual have a great influence on the results: standardized residuals with zero corrections tend to lower the estimates.

Table 10. The upper confidence limits for the total reserve and for the mean of the total reserve by the Poisson model and the Anscombe residuals.

The Quantile		**The Adjustment Type**		
		Without Corrections	**Zero-Correction**	**Zero-Correction & Stand.**
90%	Reserve	15,903,554	16,189,247	15,617,861
	Mean	15,615,874	15,655,784	15,286,299
95%	Reserve	17,427,248	17,332,017	16,760,632
	Mean	16,246,288	16,297,578	15,822,732
99%	Reserve	18,094,816	18,285,278	17,143,460
	Mean	17,429,394	17,502,041	16,829,463

5. Comparative Analysis with the Schedule P Database

In the previous section, we carried out a case study with different reserving methods in the bootstrap framework where we assessed the impact of the considered predictive models and residuals. Apart from the analytical perspective of the methods, it is also essential to compare and rank competing forecasting methods. One of the main weak points of the comparative studies in the published actuarial science-related papers is the lack of actual knowledge for which considered model is the most precise. In most of the comparative studies, to our knowledge, the chain-ladder mean is often kept as a benchmark, but in the end, this should not be the only requirement when deciding which model is the best (or the most precise). There are enough statistical tools and methods to measure the prediction accuracy.

5.1. Schedule P Database

We apply the defined models, residuals and their adjustments to the run-off triangles from practice. We extracted 10 real datasets from the Schedule P – Analysis of Losses and Loss Expenses in the National Association of Insurance Commissioners (NAIC) database, which is available on the website of the Casualty Actuarial Society. The data include major personal and commercial lines of business from U.S. property and casualty insurers. The database contains data on six lines of business, and we chose to use the workers' compensation. The triangle data correspond to the claims of the accident years 1988–1997 with a 10-year development lag. Not all of the datasets there were applicable; some of them contained too many negative values in the upper triangle, which lead to a problem in the parameter estimation procedure with the given models, and many triangles contained a high number of zeros both in the upper and the lower triangle. Thus, we had to carefully extract the datasets, which would fulfill the requirements of the models' assumptions. Both upper and lower triangles are included, so that we can use the data to test the models' performance retrospectively, i.e., the validation process is based on the back-testing idea, and all of the methods provide reserve estimates by predicting in the same lower triangle. We used the full triangles with the identifiers 337; 1767; 2135; 2712; 7080; 8672; 34,576; 21,172; 18,767; 14,176. Anyone interested could easily find these chosen datasets from the corresponding website.

5.2. Model Validation

This subsection describes the validation process for the three methods discussed in Section 2 in combination with the possible residual definitions in the bootstrap procedure discussed in Section 3. We consider the scoring rule to measure the accuracy of probabilistic predictions. There are many scoring rules available to apply, including entire parametrized families of proper scoring rules. In accordance with the [30] prequential principle, the evaluation of probabilistic forecasts is required to be based only on the predictive distributions and the observations.

Scoring rules provide summary measures of predictive performances, by assigning numerical scores to the probabilistic forecasts and on the value that materializes. Sharpness and calibration are combined here in one measure. Sharpness refers to the concentration of the predictive distributions and is a property of the forecasts only. The less variability in the predictions, the more concentrated the predictive distributions are. Consequently, forecasts will be more sharper, and subject to the calibration, the sharper the forecasts are the better. Following [31], we denote $s(P, x)$ as the assigned score for the issued predictive distribution P and materialized observation x drawn from Q. We take scores to be penalties that the forecaster wishes to minimize. A scoring rule is proper if the expected value of the penalty $s(P, x)$ for an observation x is minimized if $P = Q$. We talk about a strictly proper scoring rule if the minimum is unique. For an introduction to scoring rules, we refer to [32,33].

We consider scoring rule that depends on first and second moments only. This type of proper scoring rule was studied by [34]. In this paper, we use the Dawid–Sebastiani scoring rule (DSS), which is defined as:

$$DSS = \left(\frac{x - \mu_P}{\sigma_P}\right)^2 + 2\ln(\sigma_P),$$

where x is the observation that realizes, μ_P is the mean and σ_P is the standard deviation of the predictive distribution. To assess the predictive performance of each model with different residual adjustments discussed in this paper, we obtain an overall performance measure by averaging the DSS scores over all of the cells in the lower triangle and over each considered dataset. Let $k = 1, \ldots, m$ denote the number of datasets used. Then, the DDS scoring rule specifies to:

$$DSS = \frac{1}{m}\sum_{k=1}^{m}\left(\left(\sum_{i,j\in\nabla}\left(\frac{(C_{ij})^k - (\widehat{\mu}_{ij})^k}{(\widehat{\sigma}_{ij})^k}\right)^2 + 2\ln((\widehat{\sigma}_{ij})^k)\right)\right), \tag{16}$$

where $C_{ij}, \quad i,j \in \nabla$ denote the cells in the lower triangle, i.e., the observation that realizes (true observation in the lower triangle), $\widehat{\mu}_{ij}$ is the estimate of the corresponding mean obtained by the predictive model and $\widehat{\sigma}_{ij}$ is the estimate of the corresponding standard deviation obtained by bootstrapping. The first term focuses on calibration and the second term on sharpness. As the goal is to maximize the sharpness, we look for the model that would minimize the penalty.

5.3. Results

In this section, we present the model assessment results obtained by the scoring rule (16). In Table 11 below, we present the overall performance measure for the over-dispersed Poisson model, the gamma model and the log-normal model with the considered residuals adjustments, i.e., using residuals without corrections, zero correction and zero correction with taking into account the influence of the observation (i.e., using the standardized residuals). Thus, we have 18 different setups and combinations.

Table 11. Model validation using the Dawid–Sebastiani scoring rule (DSS). The three lowest scores are indicated in bold.

Model	**Residual**	**Type of Adjustment**		
		Without Corrections	**0's Corrected**	**0's Corrected & Stand.**
Poisson	Pearson	**36.7**	**33.9**	396.8
	Anscombe	46.2	**44.2**	598.4
Gamma	Pearson	112.3	106.6	256.9
	Anscombe	168.7	159.1	401.6
Log-normal	Pearson	103.1	98.2	251.6
	Anscombe	158.0	149.9	383.7

As we are interested in which model minimizes the score the most, we pointed out in bold three setups with the smallest numerical value; see Table 11. In general, we can draw five main conclusions from validating the considered models:

1. The over-dispersed Poisson model fits the data best. This confirms the results obtained in the previous section, where we obtained the smallest prediction errors precisely with the ODP model.
2. The gamma model and the log-normal model are behaving rather similarly, but the log-normal model is fitting the data slightly better.
3. The lowest values of the measures are obtained with the zero-corrected residuals and with the non-corrected residuals.
4. The smallest score was obtained by the zero correction with the Pearson residuals.
5. If comparing just the choice of residuals, we see that the Anscombe residuals perform more poorly than the Pearson residuals.

In Section 4, the results showed that the lowest prediction errors were obtained with the Anscombe residual; thus, we can suspect that the Anscombe residual suffered from a considerable underestimation. The highest values of the measures are obtained strictly with the standardized residuals adjusted by zero correction. Recall that in the previous section with the given particular dataset, we saw that the lowest prediction errors were obtained by the zero-corrected and standardized residuals. This is a good example to show that in the comparative study, the focus should not be only on which model gives the lowest errors, but which method actually fits the data the best. An the actuary has to be always ready to use his/her own expertise and experience in addition to well-known or most-used models in the estimation problems, as every dataset is different, we considered 18 different setups in the model validation, and as we wanted to rank the models, then we rank the first three models based on the used scoring rule: the ODP model with the Pearson residual without the correction; the ODP model with the Pearson residuals with the zero correction; and the ODP model with the Anscombe residual with the zero correction. This also shows that in some situations, the Anscombe residuals could be considered as an alternative to the Pearson residuals.

According to this case study and comparative analysis, we can say that given the obtained results in Sections 4 and 5, the method that gives the lowest prediction errors should not be confused with being the best model. Like in our case study, methods that result in the lowest variability may be, for instance, strongly suffering from the underestimation and do not fit the data after all. We considered only one scoring rule, but more investigation is required in the model validation part.

6. Discussion

In this paper, we studied the impact of the methods and the residuals on the reserve estimates and their predictive distributions. Caution is necessary when dealing with the latest development periods of the earlier accident years. The residuals of the tail are often volatile, and adjustment is hence required if they are used in the bootstrapping process. Therefore, we implemented and compared the (over-dispersed) Poisson, the gamma and the log-normal distributions in combination with the residual adjustments in the bootstrapping framework. We saw that the (over-dispersed) Poisson model and the gamma model tend to give similar point estimates, as expected, but there are bigger differences in the estimates of the prediction errors. In our case study (Section 4), we obtained the smallest errors using standardized residuals with the zero correction, but in the model validation (Section 5), we saw the methods performing somewhat contrary to the results obtained in the case study. In general, based on the case study and the model validation part, we could conclude that the over-dispersed Poisson model with the Pearson residual fits the data the best, and it looks that the best option would be to consider zero correction for the residuals. The Poisson model with the Pearson residuals appears to be a good choice in the sense that it yields the most reliable results, based on the scoring rule, whereas the Anscombe residuals with the zero correction and standardization leading to the lowest prediction errors in the first case study showed the poorest fit with the data in the model assessment section.

We can conclude that there are many different possibilities that we have to take into account before applying the bootstrap method as the prediction errors obtained by using different combinations of possible options are quite different. It is up to an actuary which result should be taken into account when making decisions in setting up the fund for reserves. The choice of a particular model remains the main struggle, but based on our research, we can draw the following conclusions:

- The large fluctuation of the values in the data substantiates the use of the over-dispersed Poisson model. The gamma model and the ODP model tend to give similar point estimates, whereas the log-normal model produces the smallest estimated claim reserves. Here, the expertise of an actuary would help to finalize a decision in model selection, depending on the company's balance of hazard and conservatism.
- When the emphasis is on prediction errors, then the ODP model should be used for the lowest prediction errors. The log-normal model tends to give irrationally high errors.
- The choice of residuals matters in bootstrapping. The Pearson residual should be preferred, but in some cases (see Table 11), the Anscombe residual could be considered.
- The adjustment of residuals is not less important than the choice of the residual; the most precise predictions are obtained with either zero-corrected residuals or without any corrections.
- The proposed model validation and assessment ideas are generic and do not depend on a particular dataset, thus constituting a useful tool in reserve estimation.

The analysis between the estimates and the actual future payments has to be carried out by the expert in the long run, in order to validate the functionality of a reserving method and identify any needed modifications. It is contended in [8] that the effectiveness of a particular reserving method and modeling can be completely tested only with an extensive case study with data from various lines of business and companies. Then, the estimated results are compared with how the claims develop over time, and only then, we can get closer to the best choice of the reserving models. Comparative studies could have a higher value when the model validation is included in the analysis. The model assessment should become a default procedure when deciding on (reserving) models. In this paper, we considered only one scoring rule, but other statistical approaches for the model assessment could be considered in the future.

Acknowledgments: The authors would like to thank Katrien Antonio for insights and for introducing us to the Schedule P database and Roel Verbelen for constructive discussions, comments and references. The authors would also like to thank the editors and three anonymous reviewers for their helpful and constructive comments that greatly contributed to improving the final version of the paper. The research was supported by institutional research funding IUT34-5 of the Estonian Ministry of Education and Research.

Author Contributions: The main ideas of this research article were developed jointly between L.T.and M.K. L.T. and R.V. analyzed the data and were responsible for the implementation. L.T. and M.K. performed the literature review, and L.T. wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Norberg, R. Prediction of outstanding liabilities in non-life insurance. *ASTIN Bull.* **1993**, *23*, 95–115.
2. Haastrup, S.; Arjas, E. Claims reserving in continuous time: A nonparametric Bayesian approach. *ASTIN Bull.* **1996**, *26*, 139–164.
3. Antonio, K.; Plat, R. Micro–level stochastic loss reserving for general insurance. *Scand. Actuar. J.* **2014**, *7*, 649–669.
4. Taylor, G. *Loss Reserving: An Actuarial Perspective*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2000.
5. Friedland, J. *Estimating Unpaid Claims Using Basic Techniques*; Casualty Actuarial Society: Arlington, VA, USA, 2010.
6. Martínez Miranda, M.D.; Nielsen, J.P.; Verrall, R.J. Double Chain Ladder. *ASTIN Bull.* **2012**, *42*, 59–76.
7. Martínez Miranda, M.D.; Nielsen, J.P.; Sperlich, S.; Verrall, R. Continuous Chain Ladder: Reformulating and generalizing a classical insurance problem. *Expert Syst. Appl.* **2013**, *40*, 5588–5603.

8. England, P.D.; Verrall, R.J. Stochastic claims reserving in general insurance. *Br. Actuar. J.* **2002**, *8*, 443–518.
9. Kaas, R.; Goovaerts, M.; Dhaene, J.; Denuit, M. *Modern Actuarial Risk Theory: Using R*; Springer: Berlin/Heidelberg, Germany, 2008.
10. Wüthrich, M.V.; Merz, M. Stochastic Claims Reserving Methods in Insurance. In *Wiley Finance*; John Wiley & Sons: New York, NY, USA, 2008; Volume 435.
11. Renshaw, A.E.; Verrall, R.J. A stochastic model underlying the chain-ladder technique. *Br. Actuar. J.* **1998**, *4*, 903–923.
12. Mack, T. A simple parametric model for rating automobile insurance or estimating IBNR claims reserves. *ASTIN Bull.* **1991**, *21*, 93–109.
13. Verrall, R.J. An investigation into stochastic claims reserving models and the chain-ladder technique. *Insur. Math. Econ.* **2000**, *26*, 91–99.
14. Kremer, E. IBNR-claims and the two-way model of ANOVA. *Scand. Actuar. J.* **1982**, *1982*, 47–55.
15. Mack, T. Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bull.* **1993**, *23*, 213–225.
16. Taylor, G. Claims triangles/loss reserves. *Predict. Model. Appl. Actuar. Sci.* **2014**, *1*, 449.
17. England, P.; Verrall, R. Analytic and bootstrap estimates of prediction errors in claims reserving. *Insur. Math. Econ.* **1999**, *25*, 281–293.
18. Ashe, F. An essay at measuring the variance of estimates of outstanding claim payments. *ASTIN Bull.* **1986**, *16*, S99–S113.
19. Pinheiro, P.J.; Andrade e Silva, J.M.; de Lourdes Centeno, M. Bootstrap methodology in claim reserving. *J. Risk Insur.* **2003**, *70*, 701–714.
20. McCullagh, P.; Nelder, J. *Generalized Linear Models*, 2nd ed.; Chapman and Hall: New York, NY, USA, 1989.
21. De Jong, P.; Heller, G.Z. *Generalized Linear Models for Insurance Data*; Cambridge University Press: Cambridge, UK, 2008.
22. Zehnwirth, B. Probabilistic development factor models with applications to loss reserve variability, prediction intervals and risk based capital. *Casualty Actuar. Soc. Forum* **1994**, *2*, 447–606.
23. Kuang, D.; Nielsen, B.; Nielsen, J. Identification of the age-period-cohort model and the extended chain-ladder model. *Biometrika* **2008**, *95*, 979–986.
24. Kuang, D.; Nielsen, B.; Nielsen, J.P. Forecasting with the age-period-cohort model and the extended chain-ladder model. *Biometrika* **2008**, *95*, 987–991.
25. Kuang, D.; Nielsen, B.; Nielsen, J.P. Chain-ladder as maximum likelihood revisited. *Ann. Actuar. Sci.* **2009**, *4*, 105–121.
26. Schmidt, K.D. A note on the overdispersed Poisson family. *Insur. Math. Econ.* **2002**, *30*, 21–25.
27. Kuang, D.; Nielsen, B.; Nielsen, J. The geometric chain-ladder. *Scand. Actuar. J.* **2015**, *2015*, 278–300.
28. Renshaw, A.E. *On the Second Moment Properties and the Implementation of Certain GLIM Based Stochastic Claims Reserving Models*; Department of Actuarial Science & Statistics, City University: Hong Kong, China, 1994.
29. Davison, A.C.; Hinkley, D.V. *Bootstrap Methods and Their Application*; Cambridge University Press: Cambridge, UK, 1997; Volume 1.
30. Dawid, A.P. Present position and potential developments: Some personal views: Statistical theory: The prequential approach. *J. R. Stat. Soc. Ser. A (Gen.)* **1984**, *147*, 278–292.
31. Gneiting, T.; Balabdaoui, F.; Raftery, A.E. Probabilistic forecasts, calibration and sharpness. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2007**, *69*, 243–268.
32. Gneiting, T.; Raftery, A. Strictly proper scoring rules, prediction and estimation. *J. Am. Stat. Assoc.* **2007**, *102*, 359–378.
33. Riebler, A.; Held, L.; Rue, H. Estimation and extrapolation of time trends in registry data—Borrowing strength from related populations. *Ann. Appl. Stat.* **2012**, *6*, 304–333.
34. Dawid, A.P.; Sebastiani, P. Coherent dispersion criteria for optimal experimental design. *Ann. Stat.* **1999**, *27*, 65–81.