Scandinavian Actuarial Journal

ISSN: 0346-1238 (Print) 1651-2030 (Online) Journal homepage: http://www.tandfonline.com/loi/sact20

Non-parametric and parametric bootstrap techniques for age-to-age development factor methods in stochastic claims reserving

Susanna Björkwall , Ola Hössjer & Esbjörn Ohlsson

To cite this article: Susanna Björkwall , Ola Hössjer & Esbjörn Ohlsson (2009) Non-parametric and parametric bootstrap techniques for age-to-age development factor methods in stochastic claims reserving, Scandinavian Actuarial Journal, 2009:4, 306-331, DOI: 10.1080/03461230903239738

To link to this article: http://dx.doi.org/10.1080/03461230903239738

Scandinavian Actuarial Journal, 2009, 4, 306–331

Taylor & Francis
Taylor & Francis Group

Applied Section

Non-parametric and parametric bootstrap techniques for age-to-age development factor methods in stochastic claims reserving

SUSANNA BJÖRKWALL*, OLA HÖSSJER* and ESBJÖRN OHLSSON†

*Department of Mathematical Statistics, Stockholm University, SE-106 91 Stockholm, Sweden
†Länsförsäkringar Alliance, SE-106 50 Stockholm, Sweden

(*Accepted 15 January 2009*)

In the literature, one of the main objects of stochastic claims reserving is to find models underlying the chain-ladder method in order to analyze the variability of the outstanding claims, either analytically or by bootstrapping. In bootstrapping these models are used to find a full predictive distribution of the claims reserve, even though there is a long tradition of actuaries calculating the reserve estimate according to more complex algorithms than the chain-ladder, without explicit reference to an underlying model. In this paper we investigate existing bootstrap techniques and suggest two alternative bootstrap procedures, one non-parametric and one parametric, by which the predictive distribution of the claims reserve can be found for other age-to-age development factor methods than the chain-ladder, using some rather mild model assumptions. For illustration, the procedures are applied to three different development triangles.

Keywords: Bootstrap; Chain-ladder; Development factor method; Development triangle; Stochastic claims reserving

1. Introduction

The provision for outstanding claims – henceforth the claims reserve – is a major contributor to the total risk of an insurance company, especially for long-tailed lines of business. In order to estimate the risk that the provisions will not suffice to pay all claims in the end, the actuary's best estimate of the outstanding claims needs to be complemented by its predictive distribution; this is the ultimo perspective. For solvency control and risk management with Dynamic Financial Analysis (DFA) we are also interested in a shorter period, say the one-year risk. The reserving risk is then the risk of a negative run-off result, due to unexpectedly large claims payments, changes in inflation regime or in the discount rate in the simulated forecast year.

A well-known method for calculating the uncertainty of the claims reserve, obtained by chain-ladder, in meeting ultimate claims, or at least its mean squared error of prediction, is the one introduced by Mack (1993) and recently treated by Buchwalder *et al*. (2006) and

Corresponding author. E-mail: sbj@student.su.se

Scandinavian Actuarial Journal
ISSN 0346-1238 print/ISSN 1651-2030 online
http://www.tandf.no/saj
DOI: 10.1080/03461230903239738

Mack *et al.* (2006). Another popular method is bootstrapping, as introduced in this context by England & Verrall (1999) and England (2002). The latter method gives a full predictive distribution without further assumptions and can easily be used also for the purpose of finding the risk in the run-off result. Therefore, we focus on bootstrap methods here.

The standard statistical approach to claims reserving would be to first specify a model, then find an estimate of outstanding claims under the model, e.g. by maximum likelihood. Finally, the model could be used to find the precision of the estimate, e.g. the mean squared error of prediction or the predictive distribution.

In practice there is a long tradition of actuaries calculating reserve estimates according to complex algorithms without explicit reference to a model. The actuaries often make ad hoc adjustments of the reserving methods in order to fit the data set under analysis. The object of the research area called stochastic claims reserving has mostly been to find a model and a method of giving a measure of the precision of the actuary's best estimate *post festum*, i.e. without the possibility of changing the estimate itself. Hence, a model that would have produced the given estimate is constructed and then used in order to find the precision of the estimate. This approach can lead to unreasonable models, which cannot be used anyway if the actuary changes her reserving method.

The object of several papers on stochastic claims reserving has been to find a model under which the best estimate is the one given by the chain-ladder method; indeed, there has been a discussion of which model is underlying the chain-ladder, see in particular Mack & Venter (2000), Verrall (2000), and Verrall & England (2000). So even though the actuary did not use a model to pick her best estimate, these articles try to find one that would make her work consistent with the standard approach of statistics, to specify the model before finding the estimate. In Verrall (2000) several underlying models, which produce the same reserve estimates as the chain-ladder method, are suggested, and it is also remarked on the importance of careful examination of the assumptions of the model and how the chosen model effects the outstanding claims.

In this paper we argue that if the actuary's best estimate is given *a priori* to the stochastic claims reserving there is no reason why the model should reflect the estimate; we should better use a reasonable model that fits the data to find the precision of the estimate. Moreover, we question the need to bootstrap an underlying model with claim distributions fully specified, which happens to reproduce the actuary's best estimate. Instead, we develop a bootstrap methodology for the data with as few model assumptions as possible, which allows the actuary to change her reserving algorithm. We assume that the claims are independent with a given mean and variance function. The mean function is typically chosen as multiplicative, but more generally, we let the reserving algorithm define the mean function in the bootstrap world, using the fitted claims. Then, using the non-parametric bootstrap approach of Pinheiro *et al.* (2003), it only remains to specify the variance function of the claims, given an underlying assumption that all standardized residual claims have (approximately) the same distribution. For comparison, we also define a parametric bootstrap version of Pinheiro's approach that requires more distributional assumptions. We consider standardized as well as unstandardized prediction errors and apply the suggested bootstrap procedures to development triangles of different types.

Section 2 contains the definitions and gives an example of an age-to-age development factor method that might be used in practice. In Section 3 the non-parametric bootstrap procedure of Pinheiro *et al.* (2003) is discussed and an alternative parametric procedure is suggested, as well as bootstrap procedures, which can be used to find the predictive distribution of other age-to-age development factor methods than the chain-ladder. Furthermore, the double bootstrap is discussed and some details of the implementation of the bootstrap procedures are commented. The bootstrap procedures are compared on three development triangles in Section 4 and finally, some technical results are collected in Appendices A and B.

2. A basic model

We consider data in the form of a triangle of n incremental observations $\{C_{ij}; i, j \in \nabla\}$, where ∇ denotes the upper, observational triangle $\nabla = \{i = 1, \ldots, t; j = 1, \ldots, t - i + 1\}$ and C_{ij} is, e.g. paid claims, number of claims, claims incurred, or some other quantity of interest of origin year i in development year j, see Table 1. For the time being we discuss paid claims. The actuary's goal is then to predict the sum of the delayed claim amounts in the lower, unobserved future triangle $\{C_{ij}; i, j \in \Delta\}$, where $\Delta = \{i = 2, \ldots, t; j = t - i + 2, \ldots, t\}$, see Table 2. We write $R = \Sigma_{\Delta} C_{ij}$ for this sum, which is the outstanding claims for which the insurance company must hold a reserve.

Above we have implicitly made the common assumption that claims are settled within the t observed years. In long-tailed business we often have no origin year with finalized claims; when needed, we extend the model so that the unknown claims extend beyond t in a tail of length u, i.e. over the development years $t, t+1, \ldots, t+u$. For simplicity, we use the notation Δ for the set of unobserved claims in this case, too.

In practice, the actuary has used some method to calculate an estimate of the outstanding claims R; in statistical terminology this is rather a *prediction* of R. We assume that the method gives estimates $\hat{m}_{ij}$ of the cell expectations $m_{ij} = E(C_{ij})$ for all claims in both ∇ and Δ, and that these estimates are functions of our observations $\nabla C \doteq \{C_{ij}; i, j \in \nabla\}$ only. (We will use the notation ∇x to denote the ∇ collection of any variable x, and similar for Δx.) The estimate of outstanding claims is then $\hat{R} = \Sigma_{\Delta} \hat{m}_{ij}$. This is the case for

Table 1. The triangle ∇ of observed incremental payments.

	Development year					
Accident year	1	2	3	...	$t-1$	t
1	C_{11}	C_{12}	C_{13}	...	$C_{1,t-1}$	$C_{1,t}$
2	C_{21}	C_{22}	C_{23}	...	$C_{2,t-1}$	
3	C_{31}	C_{32}	C_{33}	...		
$\vdots$	$\vdots$	$\vdots$	$\vdots$			
$t-1$	$C_{t-1,1}$	$C_{t-1,2}$				
t	$C_{t,1}$					

Table 2. The triangle Δ of unobserved future claim costs.

	Development year					
Accident year	1	2	3	...	$t-1$	t
1						
2						$C_{2,t}$
3					$C_{3,t-1}$	$C_{3,t}$
$\vdots$					$\vdots$	$\vdots$
$t-1$			$C_{t-1,3}$	...	$C_{t-1,t-1}$	$C_{t-1,t}$
t		$C_{t,2}$	$C_{t,3}$	...	$C_{t,t-1}$	$C_{t,t}$

age-to-age development factor methods. Note in particular that we do not assume that the reserving method is based on an explicit statistical model.

Some reserving methods operate on cumulative claims $D_{ij} = \Sigma_{\ell=1}^{j} C_{i\ell}$ rather than incremental claims C_{ij}. Let $\mu_{ij} = E(D_{ij})$. Here is an example of an age-to-age development factor method that fits our scheme:

1. The chain-ladder method (see Taylor 2000), is used to produce development factors $\hat{f}_j$ that are estimates of $f_j = \mu_{i,j+1}/\mu_{ij}$, perhaps after excluding the oldest observations and/or sole outliers in ∇.
2. For $3 < j < t$, say, the $\hat{f}_j$s are smoothed by some method, say exponential smoothing, i.e. they are replaced by estimates obtained from a linear regression of $\log(\hat{f}_j - 1)$ on j. By extrapolation in the linear regression, this also yields $\hat{f}_j$ for the tail $j = t, t+1 \ldots, t+u$. The original $\hat{f}_j$s are kept for $j \leq 3$ and the smoothed ones used for all $j > 3$.
3. Now estimates $\hat{\mu}_{ij}$ for Δ are computed as in the standard chain-ladder method.
4. Estimates of $\hat{\mu}_{ij}$ for ∇ are obtained by the process of backwards recursion described in England & Verrall (1999).
5. Finally, the obtained claim values may be discounted by some interest rate curve, or inflated by assumed claims inflation. The latter of course requires that the observations were recalculated to fixed prices in the first place.

We now have an estimator $\hat{R} = h(\nabla C)$ for some possibly quite complex function h that might be specified only by an algorithm as in the example. Our primary object is to find the bootstrap estimate of the predictive distribution of $\hat{R}$.

3. Bootstrap methods

The basic idea of bootstrapping is to work with the *Bootstrap world* in order to make inference on the *Real world* see Efron & Tibshirani (1993). This is done by investigating the result of B simulations in the bootstrap world and assuming that the conclusions from these are approximately valid in the real world; this is the so-called plug-in-principle Efron & Tibshirani (1993). With the outstanding claims in consideration this means that a relation between the true outstanding claims R and its estimator $\hat{R}$ in the real world can be substituted in the bootstrap world by their bootstrap counterparts R^{**}, i.e. the true outstanding claims in the bootstrap world, and $\hat{R}^*$, i.e. the estimated outstanding claims in the bootstrap world. Hence, the process error and the estimation error are considered by R^{**} and $\hat{R}^*$, respectively. This makes it possible to approximate the mean square error of prediction as well as its predictive distribution. Henceforth we use the index '*' for random variables or plug-in estimators in the bootstrap world which correspond to observations or estimators in the real world, while the index '**' is used for random variables in the bootstrap world when the counterparts in the real world are unobserved.

Pinheiro *et al.* (2003) use the plug-in-principle to obtain the predictive distribution of R by a non-parametric bootstrap technique which is documented in a general context by Davison & Hinkley (1997). Even though Pinheiro *et al.* (2003) adopt the statistical assumptions underlying the chain-ladder in the literature, the bootstrap procedure can easily be extended to other reserving algorithms as well since the plug-in-principle is used. Hence, our purpose is to modify it to a non-parametric bootstrap procedure which works for age-to-age development factor methods used in practice, e.g. the one described in Section 2. We also suggest a completely parametric approach consistent with, and as a complement to, the non-parametric procedure.

3.1. Bootstrapping data with a generalized linear model (GLM) using standardized prediction errors

Some assumptions about the model structure of ∇C have to be imposed in order to bootstrap the data. In the literature a common choice is to use a generalized linear model (GLM), in particular an over-dispersed Poisson distribution (ODP) with a logarithmic link function. A consequence of this underlying model is that the expected claims obtained by maximum likelihood estimation of the parameters in the GLM equal the ones obtained by the chain-ladder method, if the column sums of the triangle are positive, see Renshaw & Verrall (1998). Thus, the expectations of the claims can be obtained either by maximum likelihood estimation or by the chain-ladder, while the variances, which are needed for the residuals, are given by the assumption of the GLM. Note, however, that for the results obtained by the GLM to equal the chain-ladder results the full triangle must be used. Since actuaries often use development factor methods similar to the example described in Section 2, a GLM like the ODP may be very discordant with how the actuary has actually fitted the model.

The bootstrap methods described by England & Verrall (1999), England (2002), and Pinheiro *et al.* (2003) are all based on a GLM. The method discussed in Pinheiro *et al.* (2003) assumes the following log-additive structure of the $n = t(t+1)/2$ incremental observations in ∇C:

$$\begin{aligned} &E(C_{ij}) = m_{ij} \quad \text{and} \quad \mathrm{Var}(C_{ij}) = \phi m_{ij}^{p} \\ &\log(m_{ij}) = \eta_{ij} \\ &\eta_{ij} = c + \alpha_i + \beta_j, \quad \alpha_1 = \beta_1 = 0 \end{aligned} \tag{3.1}$$

The fitted values $\nabla\hat{m}$ and the forecasts $\Delta\hat{m}$ are calculated by maximum quasi-likelihood estimation of the $q = 2t - 1$ model parameters c, α_i, and β_j, e.g. under the assumption of an ODP, i.e. $p = 1$, or a gamma distribution, i.e. $p = 2$. Estimators of the outstanding claims are then obtained by summing per accident year $\hat{R}_i = \Sigma_{j \in \Delta_i} \hat{m}_{ij}$, where Δ_i denotes the row corresponding to accident year i in $\Delta\hat{m}$. The estimator of the grand total is $\hat{R} = \Sigma_\Delta \hat{m}_{ij}$.

The residuals are needed for the resampling process and the common choice is to use the Pearson residuals

$$r_{ij}^P = \frac{C_{ij} - \hat{m}_{ij}}{\sqrt{\hat{m}_{ij}^p}}, \tag{3.2}$$

which should have approximately zero mean and constant variance. Pinheiro *et al.* (2003), as well as England & Verrall (1999) and England (2002), work under the assumption that the residuals are independent and identically distributed, an assumption that can be questioned, see e.g. Larsen (2007) and Appendix A. Nevertheless, we shall adhere to this assumption.

The Pearson residuals need to be adjusted in order to obtain (approximately) equal variance. England & Verrall (1999) and England (2002) use a global adjusting factor

$$r_{ij}^{PA} = \sqrt{\frac{n}{n-q}} r_{ij}^P, \tag{3.3}$$

whereas Pinheiro *et al.* (2003) argue that the hat matrix standardized Pearson residuals are a better choice. They are given by

$$r_{ij}^{PA} = \frac{r_{ij}^P}{\sqrt{1 - h_{ij}}}, \tag{3.4}$$

where the h_{ij}:s are the diagonal elements of the $n \times n$ hat matrix H, which for a GLM is given by

$$H = X(X^T WX)^{-1} X^T W, \tag{3.5}$$

where X is an $n \times q$ design matrix and the generic elements $W_{ij,ij}$ of the $n \times n$ diagonal matrix W are

$$W_{ij,ij} = \left(V(m_{ij}) \left(\frac{\partial \eta_{ij}}{\partial m_{ij}} \right)^2 \right)^{-1} \tag{3.6}$$

and V is the variance function.

This choice of residual correction is in accordance with Davison & Hinkley (1997). The result of the comparison in Pinheiro *et al.* (2003) does not indicate a big difference to the correction in Eq. (3.3).

Note that the residuals are also used to produce the Pearson estimate of the unknown ϕ,

$$\hat{\phi} = \frac{1}{n-q} \sum_\nabla (r_{ij}^P)^2 = \frac{1}{n} \sum_\nabla (r_{ij}^{PA})^2, \tag{3.7}$$

where the last equality is exact when Eq. (3.3) is used and an approximation for Eq. (3.4).

The next step is to get B new triangles of residuals ∇r^* by drawing samples with replacement from the collection of residuals in Eq. (3.3) or (3.4). This procedure means sampling from the empirical distribution function of the approximately independent and identically distributed residuals r.

Then B pseudo-triangles ∇C^* are generated by computing

$$C_{ij}^* = \hat{m}_{ij} + r_{ij}^* \sqrt{\hat{m}_{ij}^p} \qquad \text{for} \quad i,j \in \nabla \tag{3.8}$$

and for these B pseudo-triangles the future values $\Delta\hat{m}^*$ are forecasted by the same method as above, i.e. by estimating the parameters of the GLM. Estimators for the outstanding claims in the bootstrap world are then derived by $\hat{R}_i^* = \Sigma_{j\in\Delta_i} \hat{m}_{ij}^*$ and $\hat{R}^* = \Sigma_\Delta \hat{m}_{ij}^*$.

In order to get the random outcome of the true outstanding claims in the bootstrap world, i.e. $R_i^{**} = \Sigma_{j\in\Delta_i} C_{ij}^{**}$ and $R^{**} = \Sigma_\Delta C_{ij}^{**}$, the resampling is done once more from the empirical distribution function of the residuals to get B triangles of Δr^{**} and then computing

$$C_{ij}^{**} = \hat{m}_{ij} + r_{ij}^{**} \sqrt{\hat{m}_{ij}^p} \quad \text{for} \quad i,j \in \Delta \tag{3.9}$$

to get ΔC^{**}.

The final step is to calculate the B prediction errors and in Pinheiro *et al.* (2003) this is done by the following equations:

$$\text{pe}_i^{**} = \frac{R_i^{**} - \hat{R}_i^*}{\sqrt{\widehat{\text{Var}}(R_i^{**})}} \qquad \text{and} \qquad \text{pe}^{**} = \frac{R^{**} - \hat{R}^*}{\sqrt{\widehat{\text{Var}}(R^{**})}}. \tag{3.10}$$

The predictive distributions of the outstanding claims R_i and R are then obtained by plotting,

$$\tilde{R}_i^{**} = \hat{R}_i + \text{pe}_i^{**}\sqrt{\widehat{\text{Var}}(R_i)} \qquad \text{and} \qquad \tilde{R}^{**} = \hat{R} + \text{pe}^{**}\sqrt{\widehat{\text{Var}}(R)} \tag{3.11}$$

for each B.

We tacitly assume that the mean and variance of all bootstrapped quantities are conditional on the observed data ∇C. For instance, the variances of the bootstrapped outstanding claims are

$$\text{Var}(R_i^{**}) = \hat{\phi} \sum_{j\in\Delta_i} \hat{m}_{ij}^p \qquad \text{and} \qquad \text{Var}(R^{**}) = \hat{\phi} \sum_{\Delta} \hat{m}_{ij}^p, \tag{3.12}$$

since the variance of the bootstrapped residuals conditional on ∇C is $\hat{\phi}$ according to Eqs. (3.3), (3.4), and (3.7). Since Pinheiro *et al.* (2003), as well as England (2002), consider ϕ as constant for the data, the estimates of Eq. (3.12) appearing in Eq. (3.10) are

$$\widehat{\text{Var}}(R_i^{**}) = \hat{\phi} \sum_{j\in\Delta_i} \hat{m}_{ij}^{*p} \qquad \text{and} \qquad \widehat{\text{Var}}(R^{**}) = \hat{\phi} \sum_{\Delta} \hat{m}_{ij}^{*p} \tag{3.13}$$

and hence computable from the bootstrap world data ∇C^*. Nevertheless, ϕ is unknown and therefore

$$\widehat{\text{Var}}(R_i^{**}) = \hat{\phi}^* \sum_{j\in\Delta_i} \hat{m}_{ij}^{*p} \qquad \text{and} \qquad \widehat{\text{Var}}(R^{**}) = \hat{\phi}^* \sum_{\Delta} \hat{m}_{ij}^{*p}, \tag{3.14}$$

where

$$\hat{\phi}^* = \frac{1}{n} \sum_{\nabla} (r_{ij}^{*\ PA})^2 \tag{3.15}$$

should rather be used, see Davison & Hinkley (1997). This is in analogy with the estimated variances of the true claims reserves

$$\widehat{\text{Var}}(R_i) = \hat{\phi} \sum_{j \in \Delta_i} \hat{m}_{ij}^p \qquad \text{and} \qquad \widehat{\text{Var}}(R) = \hat{\phi} \sum_{\Delta} \hat{m}_{ij}^p, \tag{3.16}$$

which are computable from the real data ∇C, as opposed to $\text{Var}(R_i)$ and $\text{Var}(R)$.

As a complement to the non-parametric procedure described above we suggest a parametric approach. In addition to the assumptions in Eq. (3.1) we assume a full distribution F, parametrized by the mean and variance, so that we may write $F = F(m_{ij}, \phi m_{ij}^p)$. Instead of resampling the residuals, we draw $C_{ij}{}^*$ from $F(\hat{m}_{ij}, \hat{\phi}\hat{m}_{ij}^p)$ for all i, $j \in \nabla$ and thereby we directly get the pseudo-triangles ∇C^*. The bootstrap estimates $\hat{R}_i{}^* = \Sigma_{j \in \Delta_i} \hat{m}_{ij}{}^*$ and $\hat{R}^* = \Sigma_{\Delta} \hat{m}_{ij}{}^*$ are then calculated for each simulation by estimating the parameters of the GLM. In order to get $R_i{}^{**} = \Sigma_{j \in \Delta_i} C_{ij}{}^{**}$ and $R^{**} = \Sigma_{\Delta} C_{ij}{}^{**}$ we sample once again from $F(\hat{m}_{ij}, \hat{\phi}\hat{m}_{ij}^p)$ to get $C_{ij}{}^{**}$ for all i, $j \in \Delta$. Finally, the B observations of Eqs. (3.10) and (3.13) are inserted into Eq. (3.11) to yield the sought predictive distribution.

These methods of bootstrapping for claims reserve uncertainty are described in Figure 1 and are referred to as the non-parametric and the parametric standardized predictive bootstrap.

3.2. The double bootstrap

It would be preferable to use

$$\text{pe}^{**} = \frac{R^{**} - \hat{R}^*}{\sqrt{\widehat{\text{Var}}(R^{**} - \hat{R}^*)}} \tag{3.17}$$

and

$$\tilde{R}^{**} = \hat{R} + \text{pe}^{**}\sqrt{\widehat{\text{Var}}(R - \hat{R})} \tag{3.18}$$

instead of Eqs. (3.10) and (3.11), in particular if the estimation error is much larger than the process error. Although this is more complicated it can be achieved by means of a double bootstrap. However, the computational complexity of this approach is quite prohibitive because of the nested bootstrap loop and therefore the double bootstrap is not included in our numerical study.

For each of the B bootstrap replicates, we generate $\tilde{B}$ double bootstrap claims reserves R^d and estimated claims reserves $\hat{R}^d$ in analogy with R^{**} and $\hat{R}^*$ in Section 3.1, the difference being that we use ∇C^* as our data rather than ∇C. Then

$$\widehat{\text{Var}}(R - \hat{R}) = \text{Var}(R^{**} - \hat{R}^* | \nabla C) \tag{3.19}$$

and

Stage 1 - Real world

· Estimate the parameters in the GLM in equation (3.1).
· Forecast the future expected values $\Delta\hat{m}$ and calculate the fitted values $\nabla\hat{m}$.
· Calculate the residuals for the non-parametric bootstrap and $\hat{\phi}$ for the parametric bootstrap.
· Calculate the outstanding claims $\hat{R}_i = \sum_{j\in\Delta_i} \hat{m}_{ij}$ and $\hat{R} = \sum_\Delta \hat{m}_{ij}$.

Stage 2 - Bootstrap world

Substage 2.1 - The estimated outstanding claims

· Resample residuals (3.3) or (3.4) with replacement and insert into (3.8) together with $\nabla\hat{m}$ or sample from $F(\hat{m}_{ij}, \hat{\phi}\hat{m}_{ij}^p)$ for $i, j \in \nabla$ to obtain the pseudo-reality in ∇C^*.
· Estimate the parameters in the GLM with the pseudo-triangles.
· Forecast the future expected values $\Delta\hat{m}^*$.
· Calculate the estimated outstanding claims $\hat{R}_i^* = \sum_{j\in\Delta_i} \hat{m}_{ij}^*$ and $\hat{R}^* = \sum_\Delta \hat{m}_{ij}^*$.

Substage 2.2 - The true outstanding claims

· Resample residuals (3.3) or (3.4) with replacement and insert into (3.9) together with $\Delta\hat{m}$ or sample from $F(\hat{m}_{ij}, \hat{\phi}\hat{m}_{ij}^p)$ for $i, j \in \Delta$ to obtain the pseudo-reality in ΔC^{**}.
· Calculate the true outstanding claims $R_i^{**} = \sum_{j\in\Delta_i} C_{ij}^{**}$ and $R^{**} = \sum_\Delta C_{ij}^{**}$.
· Store the prediction errors pe_i^{**} and pe^{**} in (3.10).
· Return to the beginning of the bootstrap loop and repeat it B times.

Stage 3 - Analysis of the simulations

· Obtain the predictive distribution of R_i and R, the true outstanding claims in the real world, by plotting the B values in (3.11).

Figure 1. The procedure of the non-parametric and the parametric standardized predictive bootstrap.

$$\widehat{\mathrm{Var}}(R^{**} - \hat{R}^*) = \mathrm{Var}(R^d - \hat{R}^d | \nabla C^*), \tag{3.20}$$

where the last variance is approximated by the sample variance of all $\tilde{B}$ double bootstrap replicates.

An alternative to Eqs. (3.19) and (3.20) is to use the variance of the process and the estimation errors in Eq. (A.1) in Appendix B, i.e.

$$\widehat{\mathrm{Var}}(R - \hat{R}) = \widehat{\mathrm{Var}}(R) + \widehat{\mathrm{Var}}(\hat{R}) \tag{3.21}$$

and

$$\widehat{\mathrm{Var}}(R^{**} - \hat{R}^*) = \widehat{\mathrm{Var}}(R^{**}) + \widehat{\mathrm{Var}}(\hat{R}^*), \tag{3.22}$$

where the process errors are estimated by

$$\widehat{\mathrm{Var}}(R) = \hat{\phi} \sum_\Delta \hat{m}_{ij}^p \tag{3.23}$$

and

$$\widehat{\mathrm{Var}}(R^{**}) = \hat{\phi}^* \sum_\Delta \hat{m}_{ij}^{*p}. \tag{3.24}$$

The estimation errors are approximated by the sample variance of the corresponding bootstrap replicates

$$\widehat{\text{Var}}(\hat{R}) = \text{Var}(\hat{R}^*) \tag{3.25}$$

and

$$\widehat{\text{Var}}(\hat{R}^*) = \text{Var}(\hat{R}^d). \tag{3.26}$$

3.3. Bootstrapping data with a simple underlying model and a reserving algorithm using unstandardized prediction errors

For the purpose of obtaining the predictive distribution of the claims reserve by bootstrapping, the assumption of a GLM in Eq. (3.1) is unnecessarily strong. In practice the actuary seldom assumes any model for ∇C and ΔC, but only uses a reserving algorithm in order to estimate $\nabla\hat{m}$ and $\Delta\hat{m}$. Thus, when using the plug-in-principle we just need to make an assumption of the model that generates ∇C^* and ΔC^{**} from the data ∇C, while the reserving algorithm can be used in the bootstrap world too in order to estimate $\Delta\hat{m}^*$.

We follow England & Verrall (1999), England (2002), and Pinheiro et al. (2003) and assume independent claims C_{ij} and a variance function in terms of the means, i.e.

$$E(C_{ij}) = m_{ij} \quad \text{and} \quad \text{Var}(C_{ij}) = \phi\, m_{ij}^p \tag{3.27}$$

for some $p > 0$. Thus, the mean and variance of C_{ij} are still related as in Eq. (3.1), but m_{ij} need no longer satisfy the log-additive conditions in Eq. (3.1). Instead the chosen reserving algorithm implicitly specifies the structure of all m_{ij} and produces estimates of $\hat{m}_{ij}$. The bootstrap procedures are then performed as in Section 3.1 with the exception that the residuals Eq. (3.3) are used rather than Eq. (3.4). The interpretation of n and q as the number of observations and model parameters is still the same. Using the pure chain-ladder method together with the backwards recursive operation described in England & Verrall (1999) implies that $q = 2t - 1$, as for the GLM in Eq. (3.1), since this procedure demands the estimation of $t - 1$ development factors as well as the t starting values of the backwards recursive operation. Adding exponential smoothing of the development factors, like in the example in Section 2, can indeed complicate the determination of the number of model parameters. The correction factor in Eq. (3.3) can be considered as an approximation, although the number of parameters q typically depends on the amount of smoothing.

Standardized prediction errors may still be used, since Eqs. (3.10)–(3.16) continue to hold. Indeed, it is well-known that for many bootstrap procedures, resampling of standardized quantities often increases accuracy compared to using unstandardized quantities, see e.g. Hall (1995). Nevertheless, the unstandardized prediction errors

$$\text{pe}_i^{**} = R_i^{**} - \hat{R}_i^* \qquad \text{and} \qquad \text{pe}^{**} = R^{**} - \hat{R}^* \tag{3.28}$$

are useful, in particular for the purpose of studying the estimation and the process errors, but also since they are always defined. On the contrary, the denominators of Eq. (3.10) may sometimes be non-positive, yielding undefined or imaginary standardized prediction errors, see Section 3.6. The predictive distributions of the outstanding claims R_i and R are then obtained by plotting

$$\tilde{R}_i{}^{**} = \hat{R}_i + \text{pe}_i{}^{**} \qquad \text{and} \qquad \tilde{R}^{**} = \hat{R} + \text{pe}^{**} \tag{3.29}$$

for each B. These prediction errors are used in Li (2006).

The alternative bootstrap procedures discussed above are described in detail in Figure 2 and are referred to as the non-parametric and the parametric unstandardized predictive bootstrap.

3.4. A semi-parametric bootstrap approach

England & Verrall (1999) and England (2002) use other bootstrap approaches, which are described in Appendix B. In England (2002) the bootstrap counterparts of the outstanding claims in the real world are obtained by another simulation conditional on the one in Substage 2.1 in Figure 1. In this way the *process error* $R - E(R)$ is bootstrapped differently from Substage 2.2, while Substage 2.1 bootstraps the *estimation error* $\hat{R} - E(R)$. Thus, B

Stage 1 - Actuary's best estimate in the real world

· Choose an appropriate reserving method e.g. the one described in Section 2.
· Forecast the future expected values $\Delta\hat{m}$ and calculate the fitted values $\nabla\hat{m}$.
· Calculate the residuals for the non-parametric bootstrap and $\hat{\phi}$ for the parametric bootstrap.
· Calculate the outstanding claims $\hat{R}_i = \sum_{j\in\Delta_i} \hat{m}_{ij}$ and $\hat{R} = \sum_{\Delta} \hat{m}_{ij}$.

Stage 2 - Bootstrap world

Substage 2.1 - The estimated outstanding claims

· Resample residuals (3.3) with replacement and insert into (3.8) together with $\nabla\hat{m}$ or sample from $F(\hat{m}_{ij}, \hat{\phi}\hat{m}_{ij}^{p})$ for $i, j \in \nabla$ to obtain the pseudo-reality in ∇C^*.
· Apply the chosen reserving method to ∇C^* to obtain the forecasts $\Delta\hat{m}^*$.
· Calculate the estimated outstanding claims $\hat{R}_i^* = \sum_{j\in\Delta_i} \hat{m}_{ij}^*$ and $\hat{R}^* = \sum_{\Delta} \hat{m}_{ij}^*$.

Substage 2.2 - The true outstanding claims

· Resample residuals (3.3) with replacement and insert into (3.9) together with $\Delta\hat{m}$ or sample from $F(\hat{m}_{ij}, \hat{\phi}\hat{m}_{ij}^{p})$ for $i, j \in \Delta$ to obtain the pseudo-reality in ΔC^{**}.
· Calculate the true outstanding claims $R_i^{**} = \sum_{j\in\Delta_i} C_{ij}^{**}$ and $R^{**} = \sum_{\Delta} C_{ij}^{**}$.
· Store the prediction errors pe_i^{**} and pe^{**} in (3.28).
· Return to the beginning of the bootstrap loop and repeat it B times.

Stage 3 - Analysis of the simulations

· Obtain the predictive distribution of R_i and R, the true outstanding claims in the real world, by plotting the B values in (3.29).

Figure 2. The procedure of the non-parametric and the parametric unstandardized bootstrap.

triangles $\Delta\hat{m}^{\dagger}$ are obtained by sampling a random observation $\hat{m}_{ij}^{\dagger}$ from a distribution with mean $\hat{m}_{ij}{}^*$ and variance $\phi\ \hat{m}_{ij}{}^*$ for all $i, j \in \Delta$. The predictive distribution of the outstanding claims R in real world is then obtained by plotting the B values of $\tilde{R}^{\dagger} = \Sigma_{\Delta}\ m_{ij}^{\dagger}$. England (2002) suggests using, e.g. an ODP, a negative binomial or a gamma distribution as the process distribution.

In Appendix B we question if England's bootstrap technique provides us with the right predictive distribution. The reason is that the estimation error ε_2 is replaced by $-\varepsilon_2$, whereas the process error is approximately unaffected. This affects the distribution of the prediction error as soon as the distribution of ε_2 is skewed. Like the unstandardized bootstrap of Section 3.3, England's approach does not require variance estimation of the predictive distribution. In principle though, a standardized version of England's approach could be defined. As in Sections 3.1 and 3.2, the best standardization would then be obtained using double bootstrap.

England & Verrall (2006) comment on the approach of including the process error by sampling from a separate distribution, by noting that the non-parametric standardized predictive bootstrap in Pinheiro *et al.* (2003) cannot give larger extremes of the process error than the most extreme residuals observed. Nevertheless, we see no reason to assume separate distributions for the process error and the estimation error. Either we believe in the chosen distribution on the whole and use a parametric predictive bootstrap or we do not and continue to use a non-parametric predictive bootstrap.

3.5. Estimation of p

In the literature the most frequent choice of dispersion parameter is $p = 1$ in order to reproduce the chain-ladder estimates under the assumption of a GLM, but as indicated in the method example in Section 2, a pure chain-ladder is seldom used in practice. Thus, another approach would be to choose the p that best fits the data.

A straightforward way of obtaining a suitable value of p is to use the unstandardized residuals

$$r_{ij} = \sqrt{\frac{n}{n-q}}\,(C_{ij} - \hat{m}_{ij}). \tag{3.30}$$

The following relation then holds approximatively

$$E(r_{ij}^2) \approx \mathrm{Var}(C_{ij}) = \phi\ m_{ij}^p \tag{3.31}$$

and minimizing the function

$$f(p, \phi) = \sum_{i,\,j} w_{ij}\ (r_{ij}^2 - \phi\ \hat{m}_{ij}^p)^2, \tag{3.32}$$

where w_{ij} is a weight for observation C_{ij}, with respect to p and ϕ yields an estimator for p. Once a reasonable value of p is chosen and the residuals for the resampling process are defined, ϕ is estimated by Eq. (3.7). The simplest choice is to use uniform weights $w_{ij} \equiv 1$

in Eq. (3.32). Another possibility is inverse variance weighting, $w_{ij} = \widehat{\mathrm{Var}}(r_{ij}^2)^{-1}$. In order to specify these weights, further model assumptions would be needed though.

In principle, we could use non-integer values of p for the non-parametric bootstrap. However, since we focus on the comparative performance of the parametric and non-parametric bootstrap methods in this paper, we have not pursued this approach. Instead we consider the estimated value of p as an indicator of whether $p = 1$ or $p = 2$ should be used in the non-parametric bootstrap and whether an ODP or a gamma distribution should be used in the parametric bootstrap. It is of course important that this approach of estimating p is complemented by a residual analysis in order to get the relationship between the mean and the variance right as well as to detect outliers.

3.6. Implementation details

There are some major problems with the process of resampling the residuals for the non-parametric bootstrap procedures. Firstly, the bootstrap world is hardly a good approximation of the real world if the claims triangle is small. Furthermore, the basic assumption of the non-parametric bootstrap procedure of identically distributed residuals is certainly violated for $p = 1$, i.e. for an ODP, see Appendix A. Depending on the chosen reserving method and the value of p, the standardized residuals in Eq. (3.2) sometimes imply a limitation of the set of triangles that can be analyzed, since the residual will be undefined or imaginary whenever a fitted value in $\nabla\hat{m}$ is non-positive. Finally, using the residuals to solve equation Eq. (3.8) sometimes results in undesirable negative increments in the pseudo-triangles.

Thus, if the claims triangle ∇C is small, a parametric bootstrap procedure seems preferable. On the other hand, if we know nothing about F and have a large triangle, a non-parametric bootstrap procedure would be our first choice. Note, however, that a parametric bootstrap procedure does not solve the problem with undefined residuals since they are needed in order to estimate ϕ as well. Furthermore, a parametric bootstrap procedure should be used if negative increments in the pseudo-triangles are unacceptable and a gamma distribution should particularly be used if it is undesirable that the increments only take on the values zero and multiples of ϕ, which is the case for the ODP.

The choice of prediction errors causes another problem. The standardized ones in Eq. (3.10) are sensitive to pseudo-triangles where the row sums of the outstanding claims are non-positive. An ad hoc solution is simply to cut out these pseudo-triangles from the simulation process if they are rare, another solution is to use the unstandardized prediction errors in Eq. (3.28) instead. The unstandardized ones, on the other hand, result in a predictive distribution which is more skewed to the left than the distribution obtained by the standardized prediction errors, see Section 4 for more details.

Since England & Verrall (1999), England (2002), and Pinheiro *et al.* (2003) replace the maximum likelihood estimation of the parameters in Eq. (3.1) by chain-ladder when $p = 1$, the same method is adopted here for the standardized predictive distribution in Figure 1, even though the non-positive column sums of the pseudo-triangles make the estimates disagree.

4. Numerical study

The purpose of the numerical study is to compare the non-parametric and the parametric bootstrap procedures under different choices of p, F and prediction errors. Since the actuary chooses an age-to-age development factor method that fits the particular development triangle under analysis, it is difficult to find one single algorithm that works for all situations. Therefore we only use the pure chain-ladder method in the comparisons, even though the bootstrap procedures allow the use of other age-to-age development factor methods as well. From now on $B = 10{,}000$ simulations are used for each prediction. The upper 95 percent limits are studied due to higher robustness than, e.g. the 99.5 percentile, which is perhaps the most frequent choice in practice. The coefficients of variation are also presented.

4.1. The triangle from Taylor & Ashe

4.1.1. Comparison with Pinheiro *et al.* (2003). First, the well-known triangle from Taylor & Ashe (1983), called Data 1 in Table 3, is analyzed by the non-parametric standardized predictive bootstrap procedure, i.e. the bootstrap procedure described in Pinheiro *et al.* (2003). The estimated reserves and the upper 95 percent limits for $p = 1$ and $p = 2$ are presented in Table 4. The second accident year is left out from the tabulation of results when $p = 1$ since a negative increment in the northeast corner of a pseudo-triangle causes a situation with an imaginary prediction error for that year. The remaining accident years are not as sensitive to negative increments as this year.

The results of the standardized predictive bootstrap procedure are in accordance with Pinheiro *et al.* (2003). As we can see, for earlier accident years, the $p = 2$ percentiles are smaller than the $p = 1$ percentiles, whereas the opposite is true for later accident years. This is natural, since most of the future claims C_{ij} of later years have large m_{ij} and hence larger variance for $p = 2$ than for $p = 1$.

4.1.2. The choice of $\hat{\phi}$ or $\hat{\phi}^*$. We continue to use the non-parametric standardized predictive bootstrap and Data 1, but we now replace Eqs. (3.13) with (3.14) in Substage 2.2 in Figure 1. Thus, we do not consider ϕ as constant for the data and therefore we

Table 3. Data 1 from Taylor & Ashe (1983).

	1	2	3	4	5	6	7	8	9	10
1	357,848	766,940	610,542	482,940	527,326	574,398	146,342	139,950	227,229	67,948
2	352,118	884,021	933,894	1,183,289	445,745	320,996	527,804	266,172	425,046	
3	290,507	1,001,799	926,219	1,016,654	750,816	146,923	495,992	280,405		
4	310,608	1,108,250	776,189	1,562,400	272,482	352,053	206,286			
5	443,160	693,190	991,983	769,488	504,851	470,639				
6	396,132	937,085	847,498	805,037	705,960					
7	440,832	847,631	1,131,398	1,063,269						
8	359,480	1,061,648	1,443,370							
9	376,686	986,608								
10	344,014									

Table 4. The estimated reserves and the 95 percentiles of the non-parametric standardized predictive bootstrap with Eq. (3.13) used in Substage 2.2 of Figure 1 for Data 1. Chain-ladder is used for $p=1$ and maximum likelihood estimation for $p=2$.

Year	Estimated reserve	95% $p=1$	Estimated reserve	95% $p=2$
2	94,634		93,316	222,789
3	469,511	906,877	446,504	799,700
4	709,638	1,191,170	611,145	992,585
5	984,889	1,535,723	992,023	1,497,633
6	1,419,459	2,084,349	1,453,085	2,170,480
7	2,177,641	3,032,643	2,186,161	3,284,490
8	3,920,301	5,271,523	3,665,066	5,692,764
9	4,278,972	6,116,000	4,122,398	6,975,123
10	4,625,811	9,450,379	4,516,073	9,286,282
Total	18,680,856	23,616,114	18,085,772	23,033,968

replace $\hat{\phi}$ by $\hat{\phi}^*$. The results are presented in Table 5. As we can see, the replacement hardly affects the results.

Note that since $p=1$ occasionally yields $\hat{m}_{ij}^* < 0$ the corresponding Pearson residuals in the bootstrap world are imaginary while $\hat{\phi}^*$ is real. Since the assumption of an ODP for the parametric procedure occasionally yields $\hat{m}_{ij}^* = 0$, the corresponding Pearson residuals in the bootstrap world are undefined and as a result, $\hat{\phi}^*$ is undefined as well. Thus, in the sequel we use Eq. (3.13) in all simulations.

4.1.3. Maximum likelihood estimation vs. chain-ladder when $p=2$. The next step is to replace the maximum likelihood estimator of the model parameters by the chain-ladder for the non-parametric standardized predictive bootstrap when $p=2$. (We already use the chain-ladder when $p=1$, cf. Section 3.6.) Consequently, the estimated reserves in Table 6 are the same as when $p=1$ in Table 4, whereas the percentiles in Table 6 are consistently higher than in Table 4.

This is an example of bootstrapping under a model that does not produce the estimator actually employed, a model which might nevertheless be quite realistic for paid claims.

4.1.4. Non-parametric bootstrap vs. parametric bootstrap. For the purpose of comparing the non-parametric and the parametric bootstrap procedures we continue to use the

Table 5. The estimated reserves and the 95 percentiles of the non-parametric standardized predictive bootstrap when Eq. (3.13) is replaced by Eq. (3.14) in Substage 2.2 in Figure 1 for Data 1. Chain-ladder is used for $p=1$ and maximum likelihood estimation for $p=2$.

Year	Estimated reserve	95% $p=1$	Estimated reserve	95% $p=2$
2	94,634		93,316	216,698
3	469,511	889,639	446,504	796,146
4	709,638	1,186,623	611,145	978,315
5	984,889	1,533,399	992,023	1,497,722
6	1,419,459	2,082,287	1,453,085	2,136,423
7	2,177,641	3,041,716	2,186,161	3,290,061
8	3,920,301	5,290,749	3,665,066	5,738,496
9	4,278,972	6,181,331	4,122,398	6,795,927
10	4,625,811	9,328,277	4,516,073	9,476,343
Total	18,680,856	23,603,123	18,085,772	23,042,954

Table 6. The estimated reserve and the 95 percentiles of the non-parametric standardized predictive bootstrap with Eq. (3.13) used in Substage 2.2 in Figure 1 for Data 1. Chain-ladder is used for $p=2$.

Year	Estimated reserve	95% $p=2$
2	94,634	236,850
3	469,511	875,382
4	709,638	1,156,050
5	984,889	1,503,685
6	1,419,459	2,141,470
7	2,177,641	3,308,805
8	3,920,301	6,199,841
9	4,278,972	7,646,140
10	4,625,811	10,698,797
Total	18,680,856	23,991,584

standardized predictive bootstrap with chain-ladder for Data 1. See Table 7 for the upper 95 percent limits and Table 8 for the coefficients of variation, i.e. $\sqrt{\mathrm{Var}(R_i^{**})}/\hat{R}_i$ and $\sqrt{\mathrm{Var}(R^{**})}/\hat{R}$.

The results of the parametric bootstrap coincide well with the results of the non-parametric bootstrap except for the last accident year. It is well-known that the chain-ladder estimate of the outstanding claims for the last accident year is extremely sensitive to outliers in the south corner of the upper triangle. If $C_{t1}{}^*$ happens to be small in the pseudo-triangle then the corresponding reserve $\hat{R}_t{}^*$ will be small compared to $R_t{}^{**}$, which affects the prediction error in Eq. (3.10). The parametric bootstrap generates more stable $C_{t1}{}^*$s than the non-parametric bootstrap, consequently there is a discrepancy in the results of the last accident year for the non-parametric and the parametric bootstrap procedures in Tables 7 and 8. The conclusion is that the parametric bootstrap may be preferable in some cases.

4.1.5. Standardized prediction errors vs. unstandardized prediction errors. From now on the unstandardized predictive bootstrap procedures are used in all tables; the results for Data 1 are presented in Tables 9 and 10. As we can see, the percentiles for the unstandardized predictive bootstrap in Table 9 are usually lower than for the standardized

Table 7. The estimated reserve and the 95 percentiles of the non-parametric and the parametric standardized predictive bootstrap with Eq. (3.13) used in Substage 2.2 in Figures 1 and 2 for Data 1. Chain-ladder is used in both cases.

Year	Estimated reserve	Non-parametric 95% $p=1$	Parametric ODP	Non-parametric 95% $p=2$	Parametric Gamma
2	94,634			236,850	220,643
3	469,511	906,877	894,754	875,382	866,833
4	709,638	1,191,170	1,195,535	1,156,050	1,162,942
5	984,889	1,535,723	1,522,381	1,503,685	1,516,868
6	1,419,459	2,084,349	2,092,719	2,141,470	2,150,441
7	2,177,641	3,032,643	3,061,294	3,308,805	3,309,838
8	3,920,301	5,271,523	5,308,455	6,199,841	6,192,286
9	4,278,972	6,116,000	6,220,501	7,646,140	7,272,012
10	4,625,811	9,450,379	9,185,885	10,698,797	9,222,470
Total	18,680,856	23,616,114	23,606,507	23,991,584	24,095,302

Table 8. The estimated reserve and the coefficients of variation of the simulations (in %) of the non-parametric and the parametric standardized predictive bootstrap with Eq. (3.13) used in Substage 2.2 in Figures 1 and 2 for Data 1. Chain-ladder is used in both cases.

Year	Estimated reserve	Non-parametric 95% $p=1$	Parametric ODP	Non-parametric 95% $p=2$	Parametric Gamma
2	94,634			76	62
3	469,511	49	49	43	42
4	709,638	37	38	32	32
5	984,889	31	31	27	28
6	1,419,459	27	27	26	26
7	2,177,641	23	23	27	26
8	3,920,301	20	20	30	29
9	4,278,972	24	25	38	35
10	4,625,811	52	50	64	48
Total	18,680,856	16	16	15	16

predictive bootstrap in Table 7, and the same goes for the coefficients of variation. Note that there is a large discrepancy in the coefficients of variation, in Table 10, for the two choices of distribution for Year 2. The reason for the extreme values, when $p=1$ or an ODP is assumed, is discussed in Section 4.3.

In Figures 3c–d and 4c–d, the predictive distributions of the total claims reserve are plotted when assuming $p=1$ for the non-parametric bootstrap procedures and an ODP for the parametric bootstrap procedures. The predictive distribution obtained by the unstandardized bootstrap in Figure 3c is slightly skewed to the left compared to the one obtained by the standardized bootstrap in Figure 3d, which is almost symmetric. This follows since the process component (Figures 3a and 4a) has smaller variability than the estimation component (Figures 3b and 4b), and the latter is slightly skewed to the right. This skewness is to a large extent removed for the standardized prediction errors (see Eq. (3.10)), because of the denominator, but not for the unstandardized prediction errors (see Eq. (3.28)). Furthermore, from Figures 3a and 4a, it does not seem to matter in our example whether we use a non-parametric or a parametric approach for the process error, even though we agree with England & Verrall (2006) that the former choice cannot give larger extremes than the most extreme residual observed. The same holds for $p=2$ or a gamma distribution (results not shown here).

Table 9. The estimated reserve and the 95 percentiles of the non-parametric and the parametric unstandardized predictive bootstrap when chain-ladder is used for Data 1.

Year	Estimated reserve	Non-parametric 95% $p=1$	Parametric ODP	Non-parametric 95% $p=2$	Parametric Gamma
2	94,634	275,957	252,438	168,132	167,585
3	469,511	821,152	813,932	750,175	754,646
4	709,638	1,141,093	1,130,218	1,055,135	1,064,059
5	984,889	1,475,776	1,487,763	1,414,799	1,403,919
6	1,419,459	2,042,976	2,023,014	1,995,397	1,982,611
7	2,177,641	2,997,277	2,973,779	3,043,356	3,049,215
8	3,920,301	5,189,024	5,156,277	5,579,973	5,564,848
9	4,278,972	5,902,840	5,935,956	6,363,139	6,257,000
10	4,625,811	7,766,632	7,561,924	7,387,885	7,088,050
Total	18,680,856	23,197,770	23,096,637	23,109,992	23,107,180

Table 10. The estimated reserve and the coefficients of variation of the simulations (in %) of the non-parametric and the parametric unstandardized predictive bootstrap when chain-ladder is used for Data 1.

Year	Estimated reserve	Non-parametric 95% $p=1$	Parametric ODP	Non-parametric 95% $p=2$	Parametric Gamma
2	94,634	122	118	52	50
3	469,511	47	46	39	38
4	709,638	38	37	31	31
5	984,889	31	31	28	27
6	1,419,459	27	27	26	26
7	2,177,641	23	23	26	26
8	3,920,301	21	20	28	27
9	4,278,972	25	25	32	32
10	4,625,811	45	44	40	38
Total	18,680,856	16	16	17	16

4.1.6. Estimation of *p*. Estimation of p by minimizing the (unweighted) sum in Eq. (3.32) yields $p=0.7280$. Thus, $p=1$ or an ODP seems to be more reasonable for this development triangle.

4.2. A triangle of claim counts

The non-parametric and the parametric unstandardized predictive bootstrap procedures are now compared on a triangle of claim counts appearing in Taylor (2000). Because of

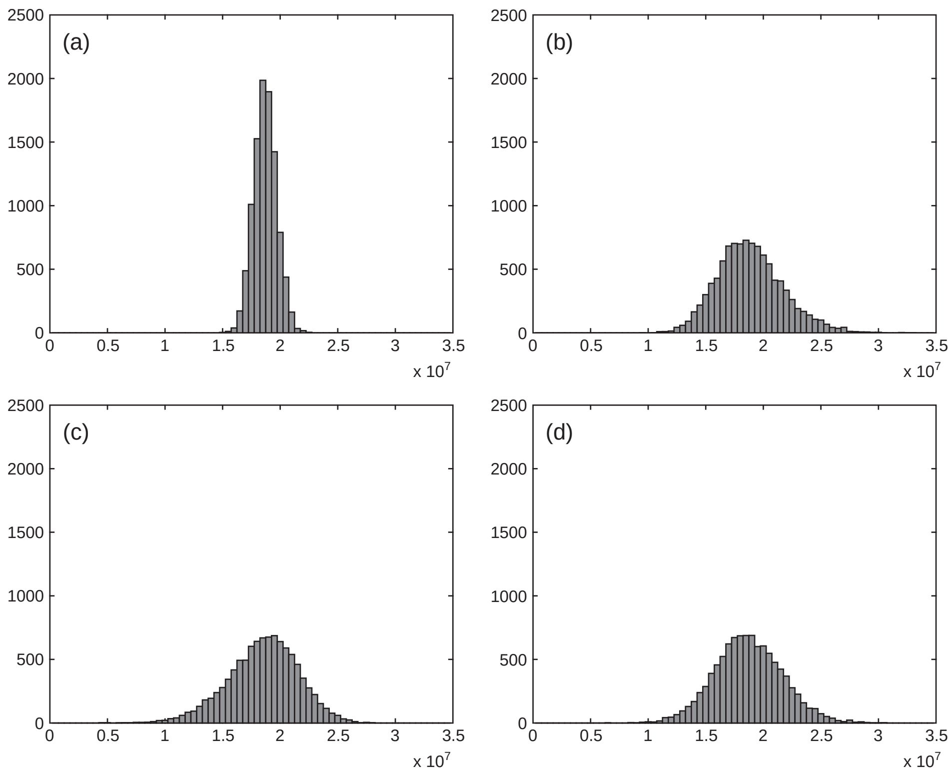

Figure 3. Density charts of (a) R^{**}, (b) $\hat{R}^{*}$, and (c) $\tilde{R}^{**}$ for the unstandardized and (d) standardized non-parametric predictive bootstrap procedures for Data 1 when $p=1$.

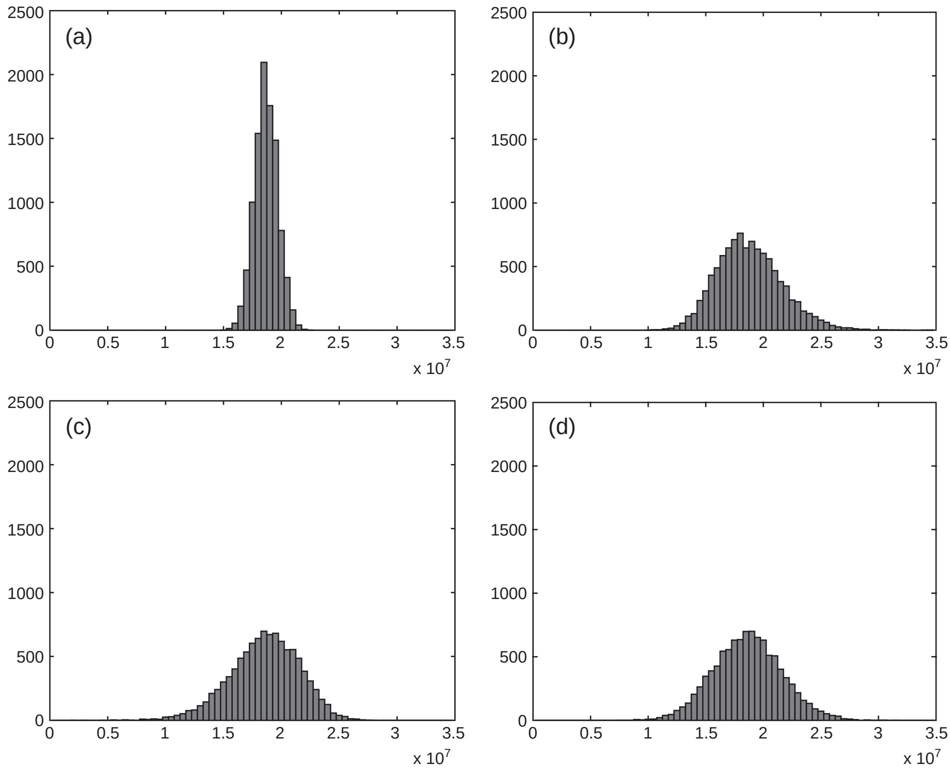

Figure 4. Density charts of (a) R^{**}, (b) $\hat{R}^{*}$, and (c) $\tilde{R}^{**}$ for the unstandardized and (d) standardized parametric predictive bootstrap procedures for Data 1 under the assumption of an ODP.

the shape of the data and in order to avoid non-positive column sums we use just the later part of the original triangle, see Table 11. This is reasonable since the claim counts from previous accident years are almost finalized.

Estimation of p yields $\hat{p} = 0.5596$, which indicates that $p = 1$ is a better choice than $p = 2$ for the non-parametric bootstrap and an ODP is preferable for the parametric bootstrap, as expected for claim counts. Nevertheless, the results for both choices are presented in Tables 12 and 13 and, as we can see, the results of the parametric bootstrap coincide well with the results of the non-parametric one.

Table 11. Data 2 from Taylor (2000).

	1	2	3	4	5	6	7
1989	589	210	29	17	12	4	9
1990	564	196	23	12	9	5	
1991	607	203	29	9	7		
1992	674	169	20	12			
1993	619	190	41				
1994	660	161					
1995	660						

Table 12. The estimated reserve and the 95 percentiles of the non-parametric and the parametric unstandardized predictive bootstrap when chain-ladder is used for Data 2.

Year	Estimated reserve	Non-parametric $p=1$	Parametric ODP	Non-parametric $p=2$	Parametric Gamma
1990	8	19	18	14	14
1991	14	26	26	20	20
1992	24	40	39	34	34
1993	36	56	55	51	50
1994	65	90	89	91	90
1995	269	323	321	400	399
Total	417	500	496	555	554

Table 13. The estimated reserve and the coefficients of variation of the simulations (in %) of the non-parametric and the parametric unstandardized predictive bootstrap when chain-ladder is used for Data 2.

Year	Estimated reserve	Non-parametric $p=1$	Parametric ODP	Non-parametric $p=2$	Parametric Gamma
1990	8	74	71	43	42
1991	14	57	55	35	33
1992	24	40	39	29	28
1993	36	32	31	26	25
1994	65	23	22	25	25
1995	269	12	12	32	31
Total	417	12	12	22	21

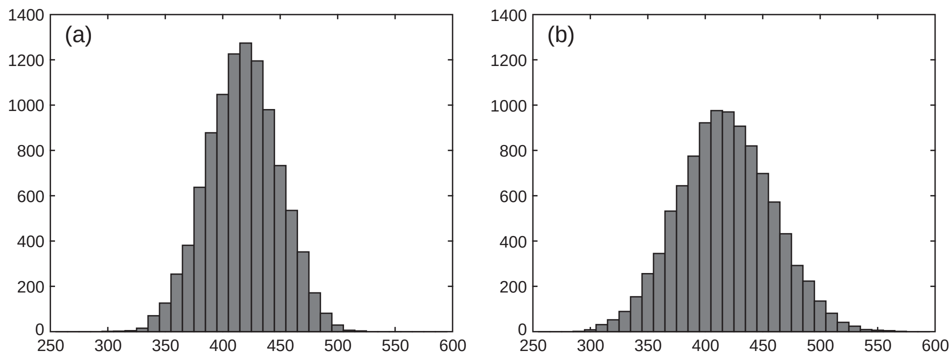

Figure 5. Density charts of (a) R^{**} and (b) $\hat{R}^{*}$ for the unstandardized non-parametric predictive bootstrap procedures for Data 2 when $p=1$.

The density charts of R^{**} and $\hat{R}^{*}$ are plotted in Figure 5. The variability of the estimation error is larger than the variability of the process error for Data 2 too, but the difference is not as extreme as for Data 1 in Figures 3 and 4.

4.3. A triangle of paid claims from a short-tailed line of business

Table 14 shows a triangle of paid claims, provided by the Swedish insurance company *AFA Försäkring*, for the short-tailed line of business *Severance Grant*.

Table 14. Data 3 provided by the Swedish insurance company AFA Försäkring.

	1	2	3	4	5	6	7
1995	48,545	56,786	32,659	12,973	4005	1696	490
1996	58,294	79,824	38,287	15,957	4617	1427	
1997	73,859	73,237	35,281	13,960	3854		
1998	65,707	67,632	32,832	12,158			
1999	92,901	80,931	36,508				
2000	66,834	47,630					
2001	45,838						

The results of the bootstrap procedures are presented in Tables 15 and 16. The percentiles for year 1996 are very different for the two choices of distribution. This is a consequence of occasional non-positive $\hat{m}_{ij}{}^*$ caused by the resampling process. Tables 17 and 18 show examples of pseudo-triangles when $p = 1$ for the non-parametric bootstrap procedure and an ODP is assumed for the parametric bootstrap procedure. By Eqs. (3.28) and (3.29) these particular simulations yield $\tilde{R}_{1996}{}^{**} = 2614$ and $\tilde{R}_{1996}{}^{**} = 2876$, respectively, which is not reasonable. Thus, even though $\hat{p} = 1.1915$, a comparison of the results for $p = 1$ and $p = 2$ indicates that $p = 2$ might be a better choice for this triangle. Another alternative might be to use a truncated ODP to exclude zero values, but this is outside the scope of the present paper.

The variability of the estimation error is larger than the variability of the process error for Data 4, but the density charts are not shown here.

5. Conclusions

So far most papers concerning bootstrapping for claims reserve uncertainty focus on obtaining the predictive distribution for the chain-ladder method by assuming underlying models, which reproduce the chain-ladder estimates. However, the assumption of an underlying model is generally not made in practice for the purpose of estimating the claims reserve, since the actuary rather uses somewhat complex reserving algorithms, without reference to statistical models. In this paper we suggest using either a non-parametric or a parametric bootstrap methodology with as few model assumptions as possible in order to make the bootstrap procedures more consistent with the actuary's way of working. The non-parametric bootstrap procedure only requires some mild distributional assumptions,

Table 15. The estimated reserve and the 95 percentiles of the non-parametric and the parametric unstandardized predictive bootstrap when chain-ladder is used for Data 3.

Year	Estimated reserve	Non-parametric $p = 1$	Parametric ODP	Non-parametric $p = 2$	Parametric Gamma
1996	621	2369	2124	873	862
1997	2408	5377	5382	3128	3116
1998	6317	10,763	10,823	8027	7960
1999	25,536	34,668	34,673	32,242	32,163
2000	46,196	59,249	58,820	58,910	58,395
2001	82,821	107,213	105,455	110,188	108,440
Total	163,898	195,586	195,097	195,876	193,573

Table 16. The estimated reserve and the coefficients of variation of the simulations (in %) of the non-parametric and the parametric unstandardized predictive bootstrap when chain-ladder is used for Data 3.

Year	Estimated reserve	Non-parametric $p = 1$	Parametric ODP	Non-parametric $p = 2$	Parametric Gamma
1996	621	173	169	26	25
1997	2408	77	74	19	18
1998	6317	44	42	17	16
1999	25,536	22	22	17	16
2000	46,196	17	17	17	17
2001	82,821	17	17	21	20
Total	163,898	12	12	12	12

Table 17. An example of pseudo-triangles when $p = 1$; the left triangle is ΔC^{**} and the right triangle is $\Delta\hat{m}^{*}$.

					757						−1236
				2007	−746					1949	−1367
			604	3300	326				2653	1561	−1095
		16,480	6116	2838	1108			17,954	3478	2046	−1435
	30,487	11,674	2527	1924	−948		26,073	11,364	2201	1295	−908
47,952	16,537	14,875	2315	640	−368	40,247	18,232	7947	1539	905	−635

Table 18. An example of pseudo-triangles when an ODP is assumed; the left triangle is ΔC^{**} and the right triangle is $\Delta\hat{m}^{*}$.

					2255						0
				1503	2255					1712	0
			3758	3006	0				4919	1716	0
		18,788	2255	3006	2255			16,892	5385	1878	0
	30,061	12,024	3006	752	752		28,929	12,041	3838	1339	0
42,085	27,055	9770	2255	2255	752	55,292	28,841	12,004	3827	1335	0

including the mean and variance function, where the actuary's choice of reserving algorithm implicitly specifies the mean structure. Consequently, the suggested bootstrap procedures can be used to obtain the predictive distribution of other age-to-age development factor methods than the chain-ladder. The non-parametric and the parametric bootstrap procedures are compared to techniques described in Pinheiro *et al.* (2003), as well as in England (2002), and finally they are applied to three development triangles.

We have seen that the results of the parametric standardized predictive bootstrap are consistent with the results of its non-parametric counterpart in Pinheiro *et al.* (2003). Furthermore, our simulation results are almost the same for the non-parametric and the parametric approach. Finally, the unstandardized predictive bootstrap procedures have revealed that the variability of the estimation error, when chain-ladder is used, is larger than the variability of the process error for all investigated development triangles. However, the difference is not that large for Data 2. This triangle provides more information than it seems since the claim counts can be summarized without any loss in statistical efficiency under the assumption of Poisson distributions. The relative size of the estimation and process errors is an interesting topic for future research. Since parameters corresponding to late origin and development years are hard to estimate for large as well as small triangles, it is by no means clear that estimation error should be relatively smaller for large triangles, in spite of the fact that there is more data available for a large triangle.

Since resampling of standardized quantities often increases accuracy compared to using unstandardized quantities, the standardized predictive bootstrap is in theory preferable to the unstandardized one. We have seen that the standardized case yields higher estimated risk, seemingly due to the fact that it makes the distribution more symmetric than the unstandardized case, where the predictive distribution is skewed to the left. A disadvantage of the standardized predictive bootstrap is that the denominators of Eq. (3.10) may sometimes be non-positive, yielding undefined or imaginary prediction errors. In principle, this could be corrected by the double bootstrap, which provides a better estimation of the variance since it includes the estimation error as well as the process error. Therefore, it would be interesting, in future papers, to analyze the behavior of the double bootstrap method both for simulated and real data sets.

In DFA, the movements of the claims reserve are of particular interest. The one-year run-off result is the change in the reserve during the financial year and is defined as the difference between the opening reserve at the beginning of the year and the sum of payments during the year and the closing reserve of the same portfolio at the end of the year. The simulation of the run-off result by the non-parametric and the parametric bootstrap procedures described in this paper would also be interesting in future papers.

Acknowledgements

The authors wish to thank an anonymous referee for valuable comments that significantly improved the quality of the paper. Ola Hössjer's research was financed by the Swedish Research Council, contract number 621-2005-2810.

References

Buchwalder, M., Bühlmann, H., Merz, M. & Wüthrich, M. (2006). The mean square error of prediction in the chain-ladder reserving method (Mack and Murphy revisited). *ASTIN Bulletin* **36** (2), 521–542.
Davison, A. C. & Hinkley, D. V. (1997). *Bootstrap methods and their applications.* Cambridge University Press.
Efron, B. & Tibshirani, R. J. (1993). *An introduction to the bootstrap.* NY: Chapman & Hall.
England, P. (2002). Addendum to "analytic and bootstrap estimates of prediction error in claims reserving". *Insurance: Mathematics and Economics* **31**, 461–466.
England, P. & Verrall, R. (1999). Analytic and bootstrap estimates of prediction errors in claims reserving. *Insurance: Mathematics and Economics* **25**, 281–293.
England, P. & Verrall, R. (2006). Predictive distributions of outstanding liabilities in general insurance. *Annals of Actuarial Science* **1** (II), 221–270.
Hall, P. (1995). *The bootstrap and Edgeworth expansion*. Berlin: Springer.
Larsen, C. R. (2007). An individual claims reserving model. *ASTIN Bulletin* **37** (1), 113–132.
Li, J. (2006). Comparison of stochastic reserving methods. *Australian Actuarial Journal* **12** (4), 489–569.
Mack, T. (1993). Distribution free calculation of the standard error of chain-ladder reserve estimates. *ASTIN Bulletin* **23** (2), 213–225.
Mack, T., Quarg, G. & Braun, C. (2006). The mean square error of prediction in the chain-ladder reserving method – a comment. *ASTIN Bulletin* **36** (2), 543–552.
Mack, T. & Venter, G. (2000). A comparison of stochastic models that reproduce chain-ladder reserve estimates. *Insurance: Mathematics and Economics* **26**, 101–107.
Pinheiro, P. J. R., Andrade e Silva, J. M. & Centeno, M. D. L. (2003). Bootstrap methodology in claim reserving. *The Journal of Risk and Insurance* **4**, 701–714.

Renshaw, A. E. & Verrall, R. J. (1998). A stochastic model underlying the chain-ladder technique. *British Actuarial Journal* **4** (IV), 903–923.
Taylor, G. (2000). *Loss reserving – an actuarial perspective*. Boston, MA: Kluwer Academic Press.
Taylor, G. & Ashe, F. R. (1983). Second moments of estimates of outstanding claims. *Journal of Econometrics* **23**, 37–61.
Verrall, R. (2000). An investigation into stochastic claims reserving models and the chain-ladder technique. *Insurance: Mathematics and Economics* **26**, 91–99.
Verrall, R. & England, P. (2000). Comments on: "a comparison of stochastic models that reproduce chain-ladder reserve estimates", by Mack and Venter. *Insurance: Mathematics and Economics* **26**, 109–111.

Appendix A

The basic assumption of the resampling process of the non-parametric bootstrap is independent and identically distributed residuals. We will now motivate that the model in Eq. (3.1) gives approximately identically distributed residuals r_{ij} for the majority of residuals Eq. (3.2) or (3.3) in the upper triangle (not close to any of the corners) when $p=2$ (gamma distribution), but not for $p=1$ (ODP). By large triangles we mean that $t\to\infty$ and hence also $n\to\infty$. For each *fixed ij*, $\hat{m}_{ij}$ is a consistent estimate of m_{ij} as n grows, and $q/n\to 0$. Hence, for large n, the residuals can be written as

$$r_{ij}=\frac{C_{ij}-m_{ij}}{\sqrt{m_{ij}^{p}}}.$$

Since the moment generating function of a $\Gamma(\alpha,\beta)$ distribution is $M(t)=(1-\beta t)^{-\alpha}$ and $p=2$ is equivalent to $C_{ij}\in\Gamma(\frac{1}{\phi},\ \phi\, m_{ij})$, the residuals r_{ij} are identically distributed according to

$$M_{r_{ij}}(t)=e^{-t}\ M_{C_{ij}}\left(\frac{t}{m_{ij}}\right)=e^{-t}(1-\phi\ t)^{-\frac{1}{\phi}}.$$

The moment generating function of a $\mathrm{Po}(\lambda)$ distribution is $M(t)=e^{\lambda(e^{t}-1)}$, but since $p=1$ implies an ODP we need a help variable X_{ij} in order to find the distribution of the residuals. The underlying model is fulfilled if $C_{ij}=\phi X_{ij},\ X_{ij}\in\mathrm{Po}\left(\frac{m_{ij}}{\phi}\right)$ and the residuals are distributed according to

$$M_{r_{ij}}(t)=e^{-t\sqrt{m_{ij}}}\ M_{C_{ij}}\left(\frac{t}{\sqrt{m_{ij}}}\right)=e^{-t\sqrt{m_{ij}}}\ M_{X_{ij}}\left(\frac{\phi\ t}{\sqrt{m_{ij}}}\right)=e^{-t\sqrt{m_{ij}}}\ e^{\frac{m_{ij}}{\phi}\left(e^{\frac{\phi\ t}{\sqrt{m_{ij}}}}-1\right)}.$$

The distributions of the residuals r_{ij} depend on m_{ij} and consequently the residuals cannot be identically distributed. However, for large m_{ij}, by the Central Limit Theorem, the distribution of r_{ij} is close to a normal distribution with mean 0 and variance ϕ, making the assumption of identically distributed residuals approximately valid.

Appendix B

In order to find the variability of the claims reserve obtained by the chain-ladder method, England & Verrall (1999) assume the model structure in Eq. (3.1). They find that the mean square error of prediction can be decomposed as

$$MSEP(R) \approx \mathrm{Var}(R) + \mathrm{Var}(\hat{R}) \tag{B.1}$$

and under the assumed model

$$\mathrm{Var}(R) = \sum_{\Delta} \phi\ m_{ij}^{p}, \tag{B.2}$$

which can easily be estimated analytically. Furthermore, in the ODP case when $p = 1$, Eq. (A.2) simplifies to

$$\mathrm{Var}(R) = \phi\ R. \tag{B.3}$$

England & Verrall (1999) suggest the use of bootstrapping to estimate the second term in Eq. (A.1). Hence, when $p = 1$ they replace Eq. (A.1) by

$$\widehat{MSEP}(\hat{R}) \approx \hat{\phi}\hat{R} + \widehat{\mathrm{Var}}(\hat{R}^*), \tag{B.4}$$

where $\widehat{\mathrm{Var}}(\hat{R}^*)$ is the variance of the B simulated values of $\hat{R}^*$ obtained by the non-parametric standardized bootstrap procedure in Substage 2.1 in Figure 1. However, England & Verrall (1999) substitute the maximum likelihood estimates of the model parameters in Figure 1 by the chain-ladder method.

In order to obtain a complete predictive distribution England (2002) extended the method in England & Verrall (1999) by replacing the analytic calculation of the process error by another simulation conditional on the bootstrap simulation. The process error is included to the B triangles $\Delta\hat{m}^*$ by sampling a random observation from a process distribution with mean $\hat{m}_{ij}{}^*$ and variance $\phi\ \hat{m}_{ij}{}^*$ to obtain the future claims $\Delta m^{\dagger}$. Here we denote the second bootstrap stage by a dagger '†' to distinguish from our own procedure of drawing the second bootstrap sample in the same way as the first, denoted by two asterisks '**'. England (2002) suggests using, e.g. an ODP, a negative binomial or a gamma distribution as the process distribution.

The predictive distribution of the outstanding claims is then obtained by plotting the B values of $R^{\dagger} = \Sigma_{\Delta}\ m_{ij}^{\dagger}$ and finally the standard deviation of the simulations gives the standard error of prediction of the outstanding claims.

England (2002) presents no justification of this procedure, but sampling from ODPs with mean $\hat{m}_{ij}{}^*$ and variance $\phi\ \hat{m}_{ij}{}^*$ will indeed provide us with a bootstrap standard error of prediction consistent with Eq. (A.1). Recall that we tacitly assume that all moments of bootstrapped quantities are conditional on the observed data ∇C. Since

$$E(R^{\dagger} \mid \Delta\hat{m}^*) = \sum_{\Delta} E(m_{ij}^{\dagger} \mid \Delta\hat{m}^*) = \sum_{\Delta} \hat{m}_{ij}{}^* = \hat{R}^*$$

and

$$\mathrm{Var}(R^{\dagger} \mid \Delta\hat{m}^*) = \sum_{\Delta} \mathrm{Var}(m_{ij}^{\dagger} \mid \Delta\hat{m}^*) = \sum_{\Delta} \hat{\phi}\, \hat{m}_{ij}{}^* = \hat{\phi}\, \hat{R}^*$$

the variance of the simulated predictive distribution is

$$\begin{aligned}\mathrm{Var}(R^{\dagger}) &= E[\mathrm{Var}(R^{\dagger} \mid \Delta\hat{m}^*)] + \mathrm{Var}[E(R^{\dagger} \mid \Delta\hat{m}^*)] \\ &= E(\hat{\phi}\, \hat{R}^*) + \mathrm{Var}(\hat{R}^*) = \hat{\phi}\, E(\hat{R}^*) + \mathrm{Var}(\hat{R}^*) \approx \hat{\phi}\, \hat{R} + \mathrm{Var}(\hat{R}^*),\end{aligned}$$

where, in the last step, we used $E(\hat{R}^*) \approx \hat{R}$ and Eq. (3.12). Note that this result is special for the ODP, since the variance is proportional to the expected value, and it is also true for the normal distribution, but it cannot be generalized to other exponential dispersion models. Hence, using an ODP in England's bootstrap will indeed provide us with a predictive distribution with the right mean and variance, but is it really the right distribution?

What we are looking for is the distribution of $R - \hat{R}$. According to the plug-in-principle the counterpart to $R - \hat{R}$ in the bootstrap world should be $R^{**} - \hat{R}^*$. Just for the moment, we use a normal distribution everywhere ($p = 0$). Hence, $R^{**} = \hat{R} + \varepsilon_1$, where $\varepsilon_1 \sim N(0;\sigma^2)$, and $\hat{R}^* = \hat{R} + \varepsilon_2$, where $\varepsilon_2 \sim N(0;\tau^2)$, for some σ^2 and τ^2. For simplicity let us say that σ^2 and τ^2 are known here. Consequently,

$$R^{**} - \hat{R}^* = \varepsilon_1 - \varepsilon_2, \tag{B.5}$$

so that according to Eq. (3.29), $\tilde{R}^{**} = \hat{R} + \varepsilon_1 - \varepsilon_2$ gives the predictive distribution of the claims reserve when the unstandardized prediction errors are used. Comparing this with England's method, $R^{\dagger}$ would be drawn from $N(\hat{R}^*; \sigma^2)$, conditionally on $\Delta\hat{m}^*$, making its unconditional distribution $N(\hat{R}, \sigma^2 + \tau^2)$ equal to that of $\tilde{R}^{**}$, due to the symmetry of the normal distribution. This means that $R^{\dagger}$ has the right predictive distribution.

Assume now a parametric bootstrap with $p = 1$ and, for simplicity, $\phi = 1$. Then $C_{ij}{}^{**} \sim \mathrm{Po}(\hat{m}_{ij})$ for all $i, j \in \Delta$, implying that Eq. (A.5) still holds with process error ε_1 and estimation error ε_2. By the additive property of the Poisson distribution, ε_1 has a centered $\mathrm{Po}(\hat{R})$ distribution. What England (2002) does is to draw $m_{ij}^{\dagger}$ from $\mathrm{Po}(\hat{m}_{ij}{}^*)$, conditionally on $\Delta\hat{m}^*$. Again, by the additive property of the Poisson distribution, we find that conditionally on $\hat{R}^*$, $R^{\dagger} \sim \mathrm{Po}(\hat{R}^*)$. This implies that

$$R^{\dagger} - \hat{R} = \varepsilon_2 + \varepsilon_3, \tag{B.6}$$

where $\varepsilon_3 \sim \mathrm{Po}(\hat{R} + \varepsilon_2) - \hat{R} - \varepsilon_2$, is a mixture of centered Poisson distributions, since ε_2 is random in the bootstrap world. Comparing the right-hand sides of Eqs. (A.6) and (A.5), we find that the estimation error occurs with opposite signs in the two equations and that the process error ε_1 has been replaced by the (closely related) ε_3. The predictive distribution in England (2002) becomes even more entangled since it is suggested that, e.g. a negative binomial or a gamma distribution can be used for the process distribution instead of an ODP even though the reserving method and the bootstrap procedure completely stands on the assumption of the latter.