DEVELOPMENT PATTERN AND PREDICTION ERROR FOR THE STOCHASTIC BORNHUETTER-FERGUSON CLAIMS RESERVING METHOD

BY

ANNINA SALUZ, ALOIS GISLER AND MARIO V. WÜTHRICH

ABSTRACT

We investigate the question how the development pattern in the Bornhuetter-Ferguson method should be estimated and derive the corresponding conditional mean square error of prediction (MSEP) of the ultimate claim prediction. An estimator of this conditional MSEP in a distribution-free model was given by Mack [9], whereas in Alai et al. [2] this conditional MSEP was studied in an over-dispersed Poisson model using the chain ladder development pattern. First we consider distributional models and derive estimators (maximum likelihood) for the development pattern taking all relevant information into account. Moreover, we suggest new estimators of the correlation matrix of these estimators and new estimators of the conditional MSEP. Our findings supplement some of Mack's results. The methodology is illustrated at two numerical examples.

KEYWORDS

Claims Reserving, Bornhuetter-Ferguson Method, Mean Square Error of Prediction, Claims Development Pattern.

1. INTRODUCTION

The chain ladder (CL) and the Bornhuetter-Ferguson (BF) method [3] are still the most frequently used claims reserving methods in practice. For both methods it is assumed that there exists a development pattern γ_j, which is the same for all accident years. For the BF method there are, in addition compared to the CL method, given estimates $\hat{\mu}_i$ of the expected ultimate claim of each accident year i, and the BF reserve of accident year i is obtained by multiplying $\hat{\mu}_i$ with the 'still to come percentage' of accident year i according to the development pattern γ_j. Thus, one essential difference between the CL and the BF philosophy is that the BF method incorporates $\hat{\mu}_i$ in the reserve estimate. Often, the CL development pattern is used for the development pattern in the

Astin Bulletin 41(2), 279-313. doi: 10.2143/AST.41.2.2136979

BF method. However, this is not consistent with the BF philosophy, as the CL development pattern disregards any information contained in the $\hat{\mu}_i$'s.

Originally, both methods were established as purely pragmatic methods without an underlying stochastic model. Such a pragmatic approach allows to determine a point estimate of the claims reserve. However, in order to assess the uncertainties in this point estimate one needs an underlying stochastic model.

For the CL method Mack [8] presented such a stochastic model and a corresponding estimator for the conditional mean square error of prediction (MSEP) in 1993. For the BF method such models and estimators are comparatively new. In 2008, Mack [9] introduced a distribution-free model underlying the BF method and suggested estimators for the corresponding MSEP. Estimators of the MSEP for models related to BF were already derived earlier as, for instance, in England-Verrall [5] or in Neuhaus [10]. However, Mack's model is probably the most general distribution-free model underlying the BF method. But this generality has also a drawback, namely, that many results in Mack [9] cannot be obtained in a stringent mathematical way and that one has to resort to pragmatic considerations and approximations.

Another approach was used by Wüthrich-Merz [13] and by Alai et al. [2]. Both studied the conditional MSEP in an over-dispersed Poisson (ODP) model where the development pattern is estimated by the CL development pattern.

In this paper we investigate the question how the development pattern in the BF method should be estimated considering all relevant information and how the corresponding correlation matrix looks like. For this purpose we consider stochastic BF models with parametric model assumptions for incremental claims. In these models we are able to derive the maximum likelihood estimators (MLEs) as well as the correlation matrix of the MLEs and derive estimators of the corresponding conditional MSEP. In contrast to the usual approach in which the chain ladder development pattern is used, we implement the BF philosophy in the estimates of the development pattern by incorporating the a priori knowledge in this estimation. In our opinion the new estimates for the development pattern, the correlation matrix and the conditional MSEP should generally be used for the BF method.

Moreover, we show how the variance of the a priori estimates can be estimated from the data. Finally, we apply our methods to real data from practice.

Organisation of the paper

In the next section we introduce the notation and data structure. In Section 3 we give the basic assumptions underlying the BF method and recall the model of Mack [9]. In Section 4 we introduce three distributional models underlying the BF method and derive estimates for the conditional MSEP in these models. Conclusions and numerical examples are given in Sections 5 and 6. Technical proofs of the statements are provided in Appendix A.

2. Notation and Data Structure

We denote the cumulative claims (cumulative payments or incurred losses) in accident year $i \in \{0, \ldots, I\}$ at the end of development year $j \in \{0, \ldots, J\}$ by $C_{i,j} > 0$ and we assume that $J \leq I$. Let $X_{i,j} = C_{i,j} - C_{i,j-1}$ denote the incremental claims, where we set $C_{i,-1} = 0$. In the sequel it is also useful to define

$$X_{[k],j} = \sum_{i=0}^{k} X_{i,j}, \; 0 \leq k \leq I, \; 0 \leq j \leq J.$$

In general we denote the summation over an index starting from 0 with a square bracket. We assume that all claims are settled after development year J and therefore the total ultimate claim of accident year i is given by $C_{i,J}$. At time I we have information

$$\mathcal{D}_I = \{C_{i,j} : i + j \leq I, \; j \leq J\},$$

and our goal is to predict $\mathcal{D}_I^c = \{C_{i,j} : i + j > I, i \leq I, j \leq J\}$. The outstanding loss liabilities for accident year i at time I are given by

$$R_i = C_{i,J} - C_{i,I-i}, \; I - J + 1 \leq i \leq I, \tag{2.1}$$

and the total outstanding loss liabilities of all accident years are given by $R = \sum_{i=I-J+1}^{I} R_i$.

Remark 2.1. The 'true' outstanding loss liabilities are given by formula (2.1) only if $C_{i,j}$ denote cumulative payments. For incurred losses $C_{i,j}$ the outstanding loss liabilities are given by

$$R_i = C_{i,J} - C_{i,I-i} + C_{i,I-i} - C_{i,I-i}^{paid}, \tag{2.2}$$

with $C_{i,j}^{paid}$ denoting the cumulative payments of accident year i up to development year j. Note that the additional term $C_{i,I-i} - C_{i,I-i}^{paid}$ is observable at time I and has no impact on the claims prediction problem and uncertainty. Therefore we only consider the outstanding loss liabilities as defined in (2.1). For incurred losses, (2.1) is often referred to as IBNR.

3. BF Reserving Method

3.1. BF Reserving Method

The BF method goes back to Bornhuetter-Ferguson [3]. There are a priori estimates of the expected ultimate claim $C_{i,J}$ given by

$$\hat{\mu}_i = v_i \, \hat{q}_i, \tag{3.1}$$

where v_i denotes the premium of accident year i and where $\hat{q}_i$ is an estimate of the expected ultimate loss ratio. The BF reserve is then given by

$$\widehat{R}_i = \hat{\mu}_i (1 - \hat{\beta}_{I-i}), \tag{3.2}$$

where $1 - \hat{\beta}_{I-i}$ is the estimated still to come factor at the end of development year $I - i$. In this paper we assume that the estimates $\hat{\mu}_i$ are independent of the observations $\mathcal{D}_I$, which is for instance the case, if these estimates represent an external expert opinion or if they come from a pricing which is mainly based on common statistics of pooled industry-wide data. Note that $\hat{\mu}_i$ is identical to an a priori estimate of the ultimate claim $C_{i,J}$, but that it is not its current predictor, which at time I is given by

$$\widehat{C}_{i,J} = C_{i,I-i} + \widehat{R}_i = C_{i,I-i} + \hat{\mu}_i (1 - \hat{\beta}_{I-i}).$$

3.2. Basic Assumptions Underlying the BF Method

From the fundamental properties of the BF method we derive some basic assumptions, which an underlying model should satisfy. These assumptions are motivated in Mack [9]. The independence assumption between the current claim amount $C_{i,I-i}$ and the reserve estimate suggests that incremental claims within the same accident year i are independent. The independence between different accident years is a standard assumption which we also adopt here. Moreover, from the BF reserve estimate (3.2) it follows that a stochastic model for the BF method has to be cross-classified of the type

$$E[C_{i,j}] = \mu_i \beta_j \quad \text{or equivalently} \quad E[X_{i,j}] = \mu_i \gamma_j,$$

with $\beta_j = \sum_{k=0}^{j} \gamma_k$. Since $\mu_i \gamma_j = (\mu_i c)(\gamma_j / c)$ for any constant $c > 0$, the parameters μ_i and γ_j are only unique up to a constant factor. Without loss of generality one can therefore assume that $\gamma_0 + \ldots + \gamma_J = 1$. The sequence $(\beta_j)_j$ then denotes the (cumulative) development pattern and $(\gamma_j)_j$ is the (incremental) development pattern. Further, we assume that the a priori estimates $\hat{\mu}_i$ are unbiased for $E[C_{i,J}]$. Hence we make the following basic assumptions for a stochastic model underlying the BF method.

Model Assumptions 3.1. (Basic Underlying Assumptions)

BF1 Incremental claims $X_{i,j}$ are independent and there exist parameters $\mu_0, \ldots, \mu_I, \gamma_0, \ldots, \gamma_J$, with $\sum_{j=0}^{J} \gamma_j = 1$ such that

$$E[X_{i,j}] = \mu_i \gamma_j, \quad 0 \le i \le I \quad \text{and} \quad 0 \le j \le J.$$

BF2 The random variables $\hat{\mu}_i$ are unbiased estimates for $\mu_i = E[C_{i,J}]$.

Mack's Model: In addition to Model Assumptions 3.1 Mack [9] makes the following variance assumption

$$\mathrm{Var}(X_{i,j}) = \mu_i \sigma_j^2. \tag{3.3}$$

Remark 3.2. Note that Mack's model is a distribution-free model for the BF method. Since it is a very general model several approximations are used to estimate the prediction uncertainties.

Alai et al. Model: In addition to Model Assumptions 3.1 Alai et al. [2] assume that incremental claims $X_{i,j}$ are ODP distributed. Prediction uncertainty is then studied within a MLE and generalized linear model framework, but by using the CL development pattern.

3.3. Estimation of the Development Pattern

In order to estimate the BF reserve and its conditional MSEP we need to estimate the development pattern γ_j, $0 \le j \le J$. In practice one often uses the CL development pattern

$$\hat{\gamma}_j^{CL} = \prod_{k=j}^{J-1} \hat{f}_k^{-1} - \prod_{k=j-1}^{J-1} \hat{f}_k^{-1}, \quad \text{where} \quad \hat{f}_k = \frac{C_{[I-k-1],k+1}}{C_{[I-k-1],k}}. \tag{3.4}$$

However, the use of the CL development pattern is not consistent with the BF philosophy, as it disregards the information in the a priori estimates $\hat{\mu}_i$. If the μ_i were known then the best linear estimate of γ_j under Model Assumptions 3.1 and under the variance assumption (3.3) would be

$$\gamma_j^* = \frac{X_{[I-j],j}}{\mu_{[I-j]}},$$

with $\mu_{[k]} = \sum_{i=0}^{k} \mu_i$, $0 \le k \le I$. Therefore Mack [9] suggests initial estimates

$$\hat{\gamma}_j^* = \frac{X_{[I-j],j}}{\hat{\mu}_{[I-j]}},$$

and then applies manual smoothing such that the final estimators $\hat{\gamma}_j$ sum up to 1.

The above estimators are not the only ones consistent with the BF philosophy, and the restriction to linear estimators is possibly not optimal. An alternative estimator is obtained with the following reasoning: assume first that we are given a full rectangle. Then an obvious estimator is

$$\hat{\gamma}_j = \frac{X_{[I],j}}{C_{[I],J}}.$$

Note that the $\hat{\gamma}_j$ automatically sum up to 1. As only the upper trapezoid $\mathcal{D}_I$ is given, a natural idea is to replace the unknown $X_{i,j}$ in the lower right triangle of $\mathcal{D}_I$ by the BF predictors $\hat{\mu}_i \hat{\gamma}_j$ resulting in the following system of equations

$$\hat{\gamma}_j = \frac{\sum_{i=0}^{I-j} X_{i,j} + \sum_{i=I-j+1}^{I} \hat{\mu}_i \hat{\gamma}_j}{\sum_{i=0}^{I-J} C_{i,J} + \sum_{i=I-J+1}^{I} \left(C_{i,I-i} + \sum_{l=I-i+1}^{J} \hat{\mu}_i \hat{\gamma}_l \right)}, \; 0 \le j \le J. \tag{3.5}$$

In order to investigate different estimators and the corresponding conditional MSEPs more precisely we consider explicit BF models with parametric assumptions for incremental claims $X_{i,j}$. All the models considered in the following will satisfy Model Assumptions 3.1 and the variance assumption (3.3). Hence they are all special cases of the model considered in Mack [9]. In these distributional models we are able to derive estimates (MLEs) of the development pattern in a stringent mathematical way. Moreover, we find explicit solutions for the correlation matrices of these estimators, which are needed to calculate the MSEP of the BF reserves. As we will see, in one of the models, the above estimator (3.5) turns out to be the estimator corresponding to the MLE, if we replace the unknown μ_i by $\hat{\mu}_i$.

4. Stochastic BF Models based on Parametric Distributions

4.1. Over-dispersed Poisson Model with Constant Dispersion Parameter

4.1.1. Model

In Model Assumptions 3.1 we have specified the basic assumptions of the BF model. In the following model we assume that the $X_{i,j}$ follow an ODP distribution with constant dispersion parameter.

Model Assumptions 4.1. (Over-dispersed Poisson Model)

P1 Incremental claims $X_{i,j}$ are independent and there exist positive parameters ϕ, $\mu_0, \ldots, \mu_I$ and $\gamma_0, \ldots, \gamma_J$ with $\sum_{j=0}^{J} \gamma_j = 1$ such that $X_{i,j}/\phi \sim$ Poisson $(\mu_i \gamma_j / \phi)$. In particular, we have

$$E[X_{i,j}] = \mu_i \gamma_j,$$
$$\mathrm{Var}(X_{i,j}) = \phi \mu_i \gamma_j,$$

where ϕ is called dispersion parameter.

P2 The a priori estimates $\hat{\mu}_i$ for $\mu_i = E[C_{i,J}]$ are unbiased and independent from $X_{l,j}$ for $0 \le l \le I$, $0 \le j \le J$.

Remarks 4.2.

- In the ODP Model the incremental claims $X_{i,j}$ are required to be positive.
- Disregarding the μ_i, that is considering data $\mathcal{D}_I$ only, the MLEs for the claims reserves and the development pattern are identical to the CL reserves and the CL development pattern $\hat{\gamma}_j^{CL}$, see for instance Hachemeister-Stanard [6], Mack [7] or Schmidt-Wünsche [12].
- Alai et al. [2] used a similar model to derive an estimate for the conditional MSEP in the BF method.
- Note that we have not made any assumption on the dependence structure between the $\hat{\mu}_i$'s.

4.1.2. Estimation of the Development Pattern

We estimate the development pattern with the MLE method. In order to calculate the MLEs we first assume that the μ_i are known and then plug in the estimates $\hat{\mu}_i$. The MLEs are best calculated using Lagrange multipliers. Note that in the ODP case the log-likelihood function for $X_{i,j}$ is given by

$$\frac{1}{\phi}\left(X_{i,j}\left(\log\mu_i + \log\gamma_j\right) - \mu_i\gamma_j\right) - g(X_{i,j},\phi),$$

where $g(\cdot,\phi)$ is a normalizing function that does not depend on the parameters μ_i and γ_j.

Remark 4.3. Multiplying the log-likelihood function by ϕ, we observe that the MLEs for the γ_j's do not depend on the dispersion parameter.

Neglecting the normalizing function g we obtain the Lagrange function with Lagrange multiplier κ given by

$$\mathcal{L}_{\mathcal{D}_I}\left(\gamma_0,\ldots,\gamma_J,\kappa\right) = \sum_{i+j\le I}\left(X_{i,j}(\log\mu_i + \log\gamma_j) - \mu_i\gamma_j\right) + \kappa\left(1 - \sum_{j=0}^{J}\gamma_j\right).$$

Observe that the last term comes from the normalization requirement $\sum_{j=0}^{J}\gamma_j = 1$. For the MLEs $\tilde{\gamma}_j$ we obtain the following equations

$$\frac{\partial\mathcal{L}_{\mathcal{D}_I}}{\partial\gamma_j} = \sum_{i=0}^{I-j}\left(\frac{X_{i,j}}{\gamma_j} - \mu_i\right) - \kappa = 0,$$

and therefore we find the solutions

$$\tilde{\gamma}_j = \frac{X_{[I-j],j}}{\mu_{[I-j]} + \kappa}. \tag{4.1}$$

From the side constraint it follows that

$$\frac{\partial \mathcal{L}_{\mathcal{D}_I}}{\partial \kappa} = 1 - \sum_{j=0}^{J} \gamma_j = 0,$$

which yields in view of (4.1) an implicit equation for κ

$$1 - \sum_{j=0}^{J} \frac{X_{[I-j],j}}{\mu_{[I-j]} + \kappa} = 0. \tag{4.2}$$

From equation (4.2) it follows that κ is the root of a polynomial of degree $J+1$. Since the γ_j need to be positive in the ODP Model and since $\mu_{[I-j]}$ is decreasing in j we have the constraint $\kappa \in (-\mu_{[I-J]}, \infty)$. In this interval there is a unique solution κ by monotonicity (cf. Remarks 4.6).

Remark 4.4. We do not use the partial derivatives with respect to μ_i. This is not necessary since we already have the a priori estimates $\hat{\mu}_i$ and therefore do not want to use MLEs for the μ_i.

Then, we replace the true (unknown) μ_i by the estimates $\hat{\mu}_i$ and obtain the estimators

$$\hat{\gamma}_j = \frac{X_{[I-j],j}}{\hat{\mu}_{[I-j]} + \hat{\kappa}}, \tag{4.3}$$

where $\hat{\kappa}$ fulfills the implicit equation

$$1 - \sum_{j=0}^{J} \hat{\gamma}_j = 0. \tag{4.4}$$

Theorem 4.5. *The estimates* (4.3) *satisfy equations* (3.5).

Proof. Observe that the $\hat{\gamma}_j$'s defined in (3.5) satisfy the constraint $\sum_{j=0}^{J} \hat{\gamma}_j = 1$. Therefore it is sufficient to prove that the estimates $\hat{\gamma}_j$ defined in (4.3) satisfy

$$\frac{\hat{\gamma}_0}{\hat{\gamma}_j} = \frac{X_{[I],0}}{X_{[I-j],j} + \sum_{i=I-j+1}^{I} \hat{\mu}_i \hat{\gamma}_j}, \quad 1 \le j \le J.$$

From equation (4.3) we have $\hat{\kappa} = \frac{X_{[I-j],j}}{\hat{\gamma}_j} - \hat{\mu}_{[I-j]}$ for $0 \le j \le J$ and conclude

$$\frac{\hat{\gamma}_0}{\hat{\gamma}_j} = \left(\frac{X_{[I],0}}{\hat{\mu}_{[I]} + \hat{\kappa}} \right) \frac{1}{\hat{\gamma}_j} = \left(\frac{X_{[I],0}}{\hat{\mu}_{[I]} + \frac{X_{[I-j],j}}{\hat{\gamma}_j} - \hat{\mu}_{[I-j]}} \right) \frac{1}{\hat{\gamma}_j} = \frac{X_{[I],0}}{X_{[I-j],j} + \sum_{i=I-j+1}^{I} \hat{\mu}_i \hat{\gamma}_j}.$$

This proves the claim. □

Remarks 4.6.

- If instead of $\hat{\mu}_i$ the data of the fully developed rectangular was given, then the MLE of γ_j would be

$$\hat{\gamma}_j^{MLE} = \frac{X_{[I],j}}{C_{[I],J}},$$

and would not depend on the μ_i at all. This shows that in the ODP Model the MLEs are not linear in the observations $X_{i,j}$. Given $\mathcal{D}_I$, the μ_i's appear in the MLEs only to the extent that the entries $X_{i,j}$ in the lower right triangle are not known and filled up with the BF predictors $\hat{\mu}_i \hat{\gamma}_j$.

- From Theorem 4.5 it follows that the estimators $\hat{\gamma}_j$ obtained by replacing the unknown μ_i in the MLEs $\tilde{\gamma}_j$ by $\hat{\mu}_i$ satisfy equations (3.5). However the easiest way of finding the solution of equations (3.5) is to calculate the estimators $\hat{\gamma}_j$ by means of equation (4.3). There we only need to calculate $\hat{\kappa}$, which can be done quite easily by starting with $\hat{\kappa} = 0$, that is with the raw estimators

$$\hat{\gamma}_j^{(0)} = \frac{X_{[I-j],j}}{\hat{\mu}_{[I-j]}},$$

and then by increasing or decreasing $\hat{\kappa}$ until the side constraint

$$f(\hat{\kappa}) = \sum_{j=0}^{J} \hat{\gamma}_j = 1$$

is fulfilled. Note hereby that in the interval $(-\hat{\mu}_{[I-J]}, \infty)$ the function $f(\hat{\kappa})$ is strictly decreasing and consequently the side constraint $f(\hat{\kappa}) = 1$ has a unique solution. Of course β_j is then estimated by $\hat{\beta}_j = \sum_{k=0}^{j} \hat{\gamma}_k$, $0 \le j \le J$, and satisfies $\hat{\beta}_J = 1$.

- From the estimators (4.3) we also see how a mathematically founded smoothing from the raw estimators $\hat{\gamma}_j^{(0)}$ to the adjusted estimators $\hat{\gamma}_j$ looks like in the ODP Model. If we write

$$\hat{\gamma}_j = c_j \hat{\gamma}_j^{(0)}$$

then the 'correction factors' c_j are given by

$$c_j = \frac{\hat{\mu}_{[I-j]}}{\hat{\mu}_{[I-j]} + \hat{\kappa}}.$$

Since $\hat{\mu}_{[I-j]}$ is decreasing in j, c_j is decreasing in j for $\hat{\kappa} > 0$ and increasing for $\hat{\kappa} < 0$. For instance, if $\hat{\kappa} < 0$ then the correction factor is greater for the late development years than for the newer ones.

- Consider the effect of conservative a priori estimation. If we increase some of the a priori estimates $\hat{\mu}_i$ then $\hat{\kappa}$ is decreasing because of the side constraint (cf. (4.3)-(4.4)). We denote the new estimates by $\hat{\mu}_i^c$, $\hat{\gamma}_j^c$ and $\hat{\kappa}^c$. For the change in the denominator of $\hat{\gamma}_j$ we then have

$$\hat{\mu}^c_{[I-k]} - \hat{\mu}_{[I-k]} + \hat{\kappa}^c - \hat{\kappa} \leq \hat{\mu}^c_{[I-j]} - \hat{\mu}_{[I-j]} + \hat{\kappa}^c - \hat{\kappa}, \quad 0 \leq j < k \leq J,$$

and because of the side constraint there must be an index $0 \leq l < J$ such that for all $j \leq l < k$ we have

$$\hat{\mu}^c_{[I-k]} - \hat{\mu}_{[I-k]} + \hat{\kappa}^c - \hat{\kappa} \leq 0 \leq \hat{\mu}^c_{[I-j]} - \hat{\mu}_{[I-j]} + \hat{\kappa}^c - \hat{\kappa}.$$

It follows that $\hat{\gamma}_k^c \geq \hat{\gamma}_k$ for all $k > l$ (in the tail) and for $j \leq l$ we have $\hat{\gamma}_j^c \leq \hat{\gamma}_j$. This means that more weight is given to the tail and the effect of conservative a priori estimation is therefore reinforced by the estimates of the development pattern. Note that in the case where we only increase $\hat{\mu}_i$ with $i \leq I - J$ the estimates of the development pattern remain unchanged (cf. (3.5)).

- The quality of the BF estimators strongly depends on the quality of the a priori estimates $\hat{\mu}_i$. If $\sum_{j=0}^{J} \hat{\gamma}_j^{(0)}$ deviates much from 1 and thus $\hat{\kappa}$ deviates much from 0, then this could be an indicator that the a priori estimates might be biased. As a further check one could estimate the development pattern and the reserves by disregarding the a priori estimates $\hat{\mu}_i$. In the OPD Model with a full development triangle or trapezoid the resulting MLEs are then identical to the CL estimates. If for most accident years the CL predictors for the ultimate claims were either higher or lower than the BF predictors, this would be a strong indicator that the a priori estimates $\hat{\mu}_i$ might be biased and that one of the basic assumptions of BF is violated. Sometimes one then adjusts the a priori estimates $\hat{\mu}_i$ (see e.g. the Cape Cod method in Wüthrich-Merz [13] or Radtke-Schmidt [11], which goes back to Bühlmann-Straub [4]). There is no problem to calculate the BF reserves with 'a posteriori' adjusted $\hat{\mu}_i$. But this means that the $\hat{\mu}_i$ are no longer independent of the data $\mathcal{D}_I$, which should be taken into account for the estimation of the MSEP and is beyond the scope of this paper.
- A tail development $\hat{\gamma}_{J+1}$ can be incorporated by replacing the side constraint in the Lagrange function by $\sum_{j=0}^{J} \gamma_j = 1 - \hat{\gamma}_{J+1}$. The estimate $\hat{\gamma}_{J+1}$ needs to come from outside together with an estimated covariance structure $\widehat{\mathrm{Cov}}(\hat{\gamma}_{J+1}, \hat{\gamma}_j)$, $0 \leq j \leq J+1$, because there is no data available for the tail.

Finally, we need to estimate the dispersion parameter ϕ. As in Wüthrich-Merz [13] we use Pearson residuals to estimate ϕ. The Pearson residuals are given by

$$R_{i,j} = \frac{X_{i,j} - \mu_i \gamma_j}{\sqrt{\mu_i \gamma_j}},$$

with $E[R_{i,j}^2] = \phi$. To estimate $R_{i,j}$ we replace μ_i and γ_j by estimates, but we need to be cautious here. We want to estimate the variance of the $X_{i,j}$ and therefore we are only interested in the process variance term and not in the parameter estimation error. The uncertainty of the external estimate $\hat{\mu}_i$ is considered in a different step. As a consequence we have to estimate μ_i and γ_j from the data in order to obtain an estimate for ϕ. We use the MLE method to gain data based estimates for μ_i and γ_j. As mentioned in Remarks 4.6 in the case of ODP data in a full triangle or trapezoid the MLEs of μ_i and γ_j coincide with the CL estimates. Therefore we estimate $R_{i,j}$ with

$$\widehat{R}_{i,j} = \frac{X_{i,j} - \widehat{C}_{i,J}^{CL}\,\hat{\gamma}_j^{CL}}{\sqrt{\widehat{C}_{i,J}^{CL}\,\hat{\gamma}_j^{CL}}},$$

where the CL development pattern $\hat{\gamma}_j^{CL}$ was defined in (3.4) and where $\widehat{C}_{i,J}^{CL} = C_{i,I-i}\prod_{j=I-i}^{J-1}\hat{f}_j$. An estimate for ϕ is then given by

$$\hat{\phi} = \frac{\sum_{i+j\leq I}\widehat{R}_{i,j}^2}{|\mathcal{D}_I| - p}, \tag{4.5}$$

where $|\mathcal{D}_I|$ is the number of observations in $\mathcal{D}_I$ and where $p = I + J + 1$ is the number of estimated parameters.

In order to quantify the uncertainties in the BF predictors $\widehat{C}_{i,J}$ and $\sum_{i=I-J+1}^{I}\widehat{C}_{i,J}$ we consider the conditional MSEP. Given information $\mathcal{I}_I$, the conditional MSEP of a predictor $\widehat{X}$ of a random variable X is defined by

$$\mathrm{msep}_{X|\mathcal{I}_I}(\widehat{X}) = E[(X - \widehat{X})^2 | \mathcal{I}_I].$$

If in addition the predictor $\widehat{X}$ is $\mathcal{I}_I$-measurable we obtain

$$\mathrm{msep}_{X|\mathcal{I}_I}(\widehat{X}) = \mathrm{Var}(X|\mathcal{I}_I) + (E[X|\mathcal{I}_I] - \widehat{X})^2.$$

For the calculation of the conditional MSEP we therefore need to consider second moments and covariances of the estimates.

4.1.3. Covariance Matrix of the Estimated Development Pattern

In this section we consider the estimation of the covariance matrix of the $\hat{\gamma}_j$'s. A well-known result from statistics is that MLEs are asymptotically unbiased and multivariate normally distributed. Moreover, the asymptotic covariance matrix is given by the inverse of the Fisher information matrix. By definition, the entries of the Fisher information matrix $H(\gamma)$ are given by

$$H(\gamma)_{j,k} = E\left[\frac{\partial l_{\mathcal{D}_I}}{\partial \gamma_j}\frac{\partial l_{\mathcal{D}_I}}{\partial \gamma_k}\right], \quad 0 \leq j, k \leq J-1,$$

where $l_{\mathcal{D}_I}$ is the log-likelihood function. In the ODP case the log-likelihood function is given by

$$l_{\mathcal{D}_I}(\gamma_0, \ldots, \gamma_{J-1}) = \sum_{\substack{i+j \le I \\ j \le J-1}} \frac{1}{\phi}\Big(X_{i,j}\big(\log \mu_i + \log \gamma_j\big) - \mu_i \gamma_j\Big)$$
$$+ \sum_{i=0}^{I-J} \frac{1}{\phi}\left(X_{i,J}\left(\log \mu_i + \log\left(1 - \sum_{j=0}^{J-1} \gamma_j\right)\right) - \mu_i\left(1 - \sum_{j=0}^{J-1} \gamma_j\right)\right) + r,$$

where r contains all remaining terms, which do not depend on $\boldsymbol{\gamma} = (\gamma_0, \ldots, \gamma_{J-1})$. Note that the last line of the above equality comes from the normalizing condition $\sum_{j=0}^{J} \gamma_j = 1$. For $0 \le j \le J-1$ we obtain the diagonal elements

$$\begin{aligned} H(\boldsymbol{\gamma})_{j,j} &= E\left[\left(\frac{\partial l_{\mathcal{D}_I}}{\partial \gamma_j}\right)^2\right] = E\left[\left(\sum_{i=0}^{I-j} \frac{1}{\phi}\left(\frac{X_{i,j}}{\gamma_j} - \mu_i\right) - \sum_{i=0}^{I-J} \frac{1}{\phi}\left(\frac{X_{i,J}}{\gamma_J} - \mu_i\right)\right)^2\right] \\ &= \sum_{i=0}^{I-j} \frac{1}{\phi^2} \operatorname{Var}\left(\frac{X_{i,j}}{\gamma_j}\right) + \sum_{i=0}^{I-J} \frac{1}{\phi^2} \operatorname{Var}\left(\frac{X_{i,J}}{\gamma_J}\right) \\ &= \sum_{i=0}^{I-j} \frac{1}{\phi} \frac{\mu_i}{\gamma_j} + \sum_{i=0}^{I-J} \frac{1}{\phi} \frac{\mu_i}{\gamma_J} = \frac{\mu_{[I-j]}}{\phi \gamma_j} + \frac{\mu_{[I-J]}}{\phi \gamma_J}. \end{aligned} \tag{4.6}$$

Analogously, for $0 \le j < k \le J-1$, we have the off-diagonal terms

$$H(\boldsymbol{\gamma})_{k,j} = H(\boldsymbol{\gamma})_{j,k} = E\left[\frac{\partial l_{\mathcal{D}_I}}{\partial \gamma_j} \frac{\partial l_{\mathcal{D}_I}}{\partial \gamma_k}\right] = \frac{\mu_{[I-J]}}{\phi \gamma_J}. \tag{4.7}$$

Hence, we write the Fisher information matrix as follows

$$H(\boldsymbol{\gamma}) = \frac{\mu_{[I-J]}}{\phi \gamma_J} \begin{pmatrix} \frac{\mu_{[I]}/\gamma_0}{\mu_{[I-J]}/\gamma_J} + 1 & 1 & \cdots & 1 \\ 1 & \frac{\mu_{[I-1]}/\gamma_1}{\mu_{[I-J]}/\gamma_J} + 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & \frac{\mu_{[I-J+1]}/\gamma_{J-1}}{\mu_{[I-J]}/\gamma_J} + 1 \end{pmatrix} \in \mathbb{R}^{J \times J}.$$

For the entries of the inverse of the Fisher information matrix we obtain (see Appendix A)

$$(H(\boldsymbol{\gamma})^{-1})_{j,k} = \frac{\phi \gamma_j}{\mu_{[I-j]}}\left(\mathbb{1}_{\{j=k\}} - \frac{\gamma_k / \mu_{[I-k]}}{\sum_{l=0}^{J} \gamma_l / \mu_{[I-l]}}\right), \quad 0 \le j,\ k \le J-1. \tag{4.8}$$

We now assume that the $\hat{\mu}_i$ are close to the true μ_i and therefore the covariances $\operatorname{Cov}(\hat{\gamma}_j, \hat{\gamma}_k)$ are close to the covariances of the true MLEs $\tilde{\gamma}_j$ (cf. (4.1)). With the asymptotic MLE approximation we then obtain

$$\mathrm{Cov}(\hat{\gamma}_j, \hat{\gamma}_k) \approx (H(\gamma)^{-1})_{j,k},$$

and of course the covariances of $\hat{\gamma}_J = 1 - \sum_{j=0}^{J-1} \hat{\gamma}_j$ are obtained by linearity. Replacing the parameters γ_j, ϕ and $\mu_{[I-j]}$ by their estimates $\hat{\gamma}_j$ (see (4.3)), $\hat{\phi}$(see (4.5)) and $\hat{\mu}_{[I-j]}$ (see Model Assumptions 4.1) we arrive at the following estimates for the covariances

$$\widehat{\mathrm{Cov}}(\hat{\gamma}_j, \hat{\gamma}_k) = \frac{\hat{\phi}\hat{\gamma}_j}{\hat{\mu}_{[I-j]}} \left(1_{\{j=k\}} - \frac{\hat{\gamma}_k / \hat{\mu}_{[I-k]}}{\sum_{l=0}^{J} \hat{\gamma}_l / \hat{\mu}_{[I-l]}} \right), \quad 0 \le j, k \le J. \tag{4.9}$$

This is clear for $0 \le j$, $k \le J - 1$ and also holds for $0 \le j$, $k \le J$.
Note that

$$\gamma_j^* = \frac{X_{[I-j],j}}{\mu_{[I-j]}}$$

are uncorrelated with variance

$$\mathrm{Var}(\gamma_j^*) = \frac{\phi \gamma_j}{\mu_{[I-j]}}$$

Hence the off-diagonal elements and the second term in (4.9) come from the side-constraint that the estimators have to sum up to one.

4.1.4. Conditional MSEP

In this section we derive estimates for the conditional MSEP of the BF predictors $\widehat{C}_{i,J} = C_{i,I-i} + \hat{\mu}_i (1 - \hat{\beta}_{I-i})$ and $\sum_{i=I-J+1}^{I} \widehat{C}_{i,j}$ under Model Assumptions 4.1. Let $\hat{\boldsymbol{\mu}} = (\hat{\mu}_0, \ldots, \hat{\mu}_I)$ and let $\mathcal{I}_I = \sigma(\mathcal{D}_I, \hat{\boldsymbol{\mu}})$ denote the σ-field containing the information at time I. Due to the $\mathcal{I}_I$-measurability of $\hat{\mu}_i$ and $\hat{\beta}_{I-i}$ we have

$$\begin{aligned} \mathrm{msep}_{C_{i,J}|\mathcal{I}_I}(\widehat{C}_{i,J}) &= E\left[\left(\sum_{j=I-i+1}^{J} X_{i,j} - \hat{\mu}_i (1 - \hat{\beta}_{I-i}) \right)^2 \middle| \mathcal{I}_I \right] \\ &= E\left[\left(\sum_{j=I-i+1}^{J} X_{i,j} - \mu_i \gamma_j \right)^2 \middle| \mathcal{I}_I \right] + \left(\sum_{j=I-i+1}^{J} \mu_i \gamma_j - \hat{\mu}_i (1 - \hat{\beta}_{I-i}) \right)^2. \end{aligned}$$

With the independence of the $X_{i,j}$ and the definition $\sum_{j=I-i+1}^{J} \gamma_j = 1 - \beta_{I-i}$ we obtain

$$\mathrm{msep}_{C_{i,J}|\mathcal{I}_I}(\widehat{C}_{i,J}) = \underbrace{\sum_{j=I-i+1}^{J} \mathrm{Var}(X_{i,j})}_{\text{Process Variance (PV}_i)} + \underbrace{(\hat{\mu}_i (1 - \hat{\beta}_{I-i}) - \mu_i (1 - \beta_{I-i}))^2}_{\text{Estimation Error (EE}_i)}.$$

The process variance reflects the randomness of the $X_{i,j}$ and the estimation error describes the uncertainty in the parameter estimates.

Process Variance. For the estimation of the process variance recall that $\mathrm{Var}(X_{i,j}) = \phi \mu_i \gamma_j$ according to Model Assumptions 4.1. If we plug in the parameter estimates we obtain the estimator

$$\widehat{\mathrm{PV}}_i = \sum_{j=I-i+1}^{J} \widehat{\mathrm{Var}}(X_{i,j}) = \sum_{j=I-i+1}^{J} \hat{\phi}\hat{\mu}_i\hat{\gamma}_j = \hat{\phi}\hat{\mu}_i(1-\hat{\beta}_{I-i}).$$

Estimation Error. Given $\mathcal{I}_I$, the term $(\hat{\mu}_i(1-\hat{\beta}_{I-i}) - \mu_i(1-\beta_{I-i}))^2$ is an unknown constant. In order to estimate this constant we rewrite it as follows

$$\begin{aligned}&\left(\hat{\mu}_i(1-\hat{\beta}_{I-i}) - \mu_i(1-\beta_{I-i})\right)^2 = \left((\hat{\mu}_i-\mu_i)(1-\hat{\beta}_{I-i}) + \mu_i(\beta_{I-i}-\hat{\beta}_{I-i})\right)^2\\ &= (1-\hat{\beta}_{I-i})^2(\hat{\mu}_i-\mu_i)^2 + \mu_i^2(\beta_{I-i}-\hat{\beta}_{I-i})^2 + 2(1-\hat{\beta}_{I-i})\mu_i(\hat{\mu}_i-\mu_i)(\beta_{I-i}-\hat{\beta}_{I-i}),\end{aligned}$$

and estimate the unknown terms $(\hat{\mu}_i-\mu_i)^2$, $(\beta_{I-i}-\hat{\beta}_{I-i})^2$ and $(\hat{\mu}_i-\mu_i)(\beta_{I-i}-\hat{\beta}_{I-i})$ by taking the mean over all possible $\mathcal{I}_I$, that is, we study the fluctuation of $\hat{\mu}_i$ and $\hat{\beta}_{I-i}$ around the true values μ_i and β_{I-i}, respectively. If we neglect the dependencies of the $\hat{\mu}_i$ and the $\hat{\beta}_j$ we obtain with the unbiasedness of the $\hat{\mu}_i$ and the (approximate) asymptotic unbiasedness of the $\hat{\beta}_j$ the following estimator for the estimation error

$$\widehat{\mathrm{EE}}_i = (1-\hat{\beta}_{I-i})^2\,\widehat{\mathrm{Var}}(\hat{\mu}_i) + \hat{\mu}_i^2\,\widehat{\mathrm{Var}}(\hat{\beta}_{I-i}),$$

where $\widehat{\mathrm{Var}}(\hat{\mu}_i)$ and $\widehat{\mathrm{Var}}(\hat{\beta}_{I-i})$ are estimates for the variances $\mathrm{Var}(\hat{\mu}_i)$ and $\mathrm{Var}(\hat{\beta}_{I-i})$, respectively. The variance of the $\hat{\beta}_{I-i}$ is estimated using the Fisher information matrix

$$\widehat{\mathrm{Var}}(\hat{\beta}_{I-i}) = \sum_{0\le j,\,k\le I-i} \widehat{\mathrm{Cov}}(\hat{\gamma}_j,\hat{\gamma}_k), \tag{4.10}$$

where the covariances $\widehat{\mathrm{Cov}}(\hat{\gamma}_j,\hat{\gamma}_k)$ are given in (4.9).

It remains to estimate the variances of the a priori estimates $\hat{\mu}_i$. An actuary being able to deliver a priori estimates $\hat{\mu}_i$ is often also able to make a statement about the precision of these estimates. In this case beside the $\hat{\mu}_i$'s there are also given a priori estimates of $\mathrm{Var}(\hat{\mu}_i)$. If this is not the case we show how $\mathrm{Var}(\hat{\mu}_i)$ can be estimated from the data. To this end we assume that the $\hat{\mu}_i$ have a constant coefficient of variation $\mathrm{CoVa}(\hat{\mu}_i) \equiv c$ and that the correlation $\mathrm{Corr}(\hat{\mu}_i, \hat{\mu}_k)$ can be estimated by

$$\widehat{\mathrm{Corr}}(\hat{\mu}_i,\hat{\mu}_k) = \begin{cases}\frac{n-|i-k|}{n}, & 0\le i,k\le I \text{ and } |i-k|<n,\\ 0, & \text{else}\end{cases} \tag{4.11}$$

where $0 \le n \le I$.

Remark 4.7. In the case where the premiums v_i and the variances of the ultimate loss ratios are constant, we obtain the above formula for the correlation $\widehat{\text{Corr}}(\hat{\mu}_i, \hat{\mu}_k)$ if the a priori loss ratios are estimated with a moving average of the ultimate loss ratios over n years (cf. Appendix A). Observe that we do not assume that the a priori estimates $\hat{\mu}_i$ are independent.

In the numerical examples we will use formula (4.11) with $n = 10$, assuming that information more than 10 years ago does not contribute to the new forecast. We define the aggregated claim amount at time I by

$$C_I = \sum_{i=0}^{I-J} C_{i,J} + \sum_{i=I-J+1}^{I} C_{i,I-i}.$$

The corresponding pure risk premium is given by

$$\Pi_I = E[C_I] = \sum_{i=0}^{I-J} \mu_i + \sum_{i=I-J+1}^{I} \beta_{I-i}\mu_i. \tag{4.12}$$

Moreover, we define the observed loss ratio at time I by

$$Q_I = \frac{C_I}{\hat{\Pi}_I} = \frac{C_I}{\Pi_I}\frac{\Pi_I}{\hat{\Pi}_I},$$

where $\hat{\Pi}_I$ is obtained from (4.12) by replacing all μ_i by the a priori estimates $\hat{\mu}_i$. With a first order Taylor approximation around Π_I we obtain

$$E[\Pi_I/\hat{\Pi}_I] \approx 1 \quad \text{and} \quad \text{Var}(\Pi_I/\hat{\Pi}_I) \approx \text{CoVa}^2(\hat{\Pi}_I),$$

and therefore with the independence of $\hat{\Pi}_I$ and C_I we conclude that $E[Q_I] \approx 1$. For the variance $\text{Var}(Q_I)$ we use the following general decomposition for the product of independent random variables X and Y

$$\text{Var}(XY) = E[X]^2\,\text{Var}(Y) + E[Y]^2\,\text{Var}(X) + \text{Var}(X)\text{Var}(Y).$$

Henceforth we have with the Taylor approximation

$$\text{Var}(Q_I) \approx \text{Var}(C_I)/\Pi_I^2 + \text{CoVa}^2(\hat{\Pi}_I) + \text{Var}(C_I/\Pi_I)\text{Var}(\Pi_I/\hat{\Pi}_I).$$

Neglecting the last term and asserting positivity we arrive at the following estimator

$$\widehat{\text{CoVa}}^2(\hat{\Pi}_I) = \max\left\{0, \widehat{\text{Var}}(Q_I) - \widehat{\text{Var}}(C_I)/\widehat{\hat{\Pi}_I^2}\right\},$$

where

$$\widehat{\hat{\Pi}}_I = \sum_{i=0}^{I-J} \hat{\mu}_i + \sum_{i=I-J+1}^{I} \hat{\beta}_{I-i}\, \hat{\mu}_i,$$

$$\widehat{\mathrm{Var}}(Q_I) = (\widehat{Q}_I - 1)^2, \quad \text{with} \quad \widehat{Q}_I = C_I / \widehat{\hat{\Pi}}_I,$$

$$\widehat{\mathrm{Var}}(C_I) = \sum_{i=0}^{I-J} \hat{\mu}_i \hat{\sigma}^2_{[J]} + \sum_{i=I-J+1}^{I} \hat{\mu}_i \hat{\sigma}^2_{[I-i]}, \text{ with } \hat{\sigma}^2_{[j]} = \hat{\phi}\hat{\beta}_j, \text{ in the ODP Model.}$$

Note that if the $\hat{\mu}_i$ were fully correlated then $\mathrm{CoVa}(\hat{\mu}_i) = \mathrm{CoVa}(\hat{\Pi}_I)$. In general we obtain with the calculation provided in Appendix A

$$\mathrm{CoVa}^2(\hat{\mu}_i) = c^2$$
$$= \max\left\{0, \mathrm{CoVa}^2(\hat{\Pi}_I)\left(1 - \frac{2}{\left(\sum_{i=0}^{I} \mu_i \beta_{I-i}\right)^2} \sum_{0 \le i < k \le I} \beta_{I-i}\, \beta_{I-k}\, \mu_i \mu_k \left(1 - \mathrm{Corr}(\hat{\mu}_i, \hat{\mu}_k)\right)\right)^{-1}\right\}, \tag{4.13}$$

where we set $\beta_j = 1$ if $j > J$. The variance is then estimated by

$$\widehat{\mathrm{Var}}(\hat{\mu}_i) = \hat{c}^2 \cdot \hat{\mu}_i^2, \tag{4.14}$$

where $\hat{c}^2$ is obtained from (4.13) by replacing the unknown $\mathrm{CoVa}^2(\hat{\Pi}_I)$, $\mathrm{Corr}(\hat{\mu}_i, \hat{\mu}_k)$ and the parameters by their estimates.

Thus, we have derived the following estimate for the conditional MSEP of $\widehat{C}_{i,J}$:

Estimate 4.8. (MSEP single accident year) *Under Model Assumptions 4.1 the conditional MSEP of the BF predictor* $\widehat{C}_{i,J}$ *is estimated by*

$$\widehat{\mathrm{msep}}_{C_{i,J}|\mathcal{I}_I}(\widehat{C}_{i,J}) = \sum_{j=I-i+1}^{J} \hat{\phi}\, \hat{\mu}_i\, \hat{\gamma}_j + \widehat{\mathrm{Var}}(\hat{\mu}_i)(1 - \hat{\beta}_{I-i})^2 + \hat{\mu}_i^2\, \widehat{\mathrm{Var}}(\hat{\beta}_{I-i}),$$

where $\widehat{\mathrm{Var}}(\hat{\beta}_{I-i})$ *is given in* (4.10) *and* $\widehat{\mathrm{Var}}(\hat{\mu}_i)$ *is given in* (4.14). *For the estimates* $\hat{\mu}_i$, $\hat{\gamma}_j$ *and* $\hat{\phi}$ *see Model Assumptions 4.1, equation* (4.3) *and* (4.5)*, respectively.*

Our next goal is the derivation of an estimate of the conditional MSEP for aggregated accident years. We have

$$\mathrm{msep}_{\sum_{i=I-J+1}^{I} C_{i,J}|\mathcal{I}_I}\left(\sum_{i=I-J+1}^{I} \widehat{C}_{i,J}\right) = E\left[\left(\sum_{i=I-J+1}^{I} C_{i,J} - \sum_{i=I-J+1}^{I} \widehat{C}_{i,J}\right)^2 \middle| \mathcal{I}_I\right]$$
$$= \sum_{i=I-J+1}^{I} \mathrm{msep}_{C_{i,J}|\mathcal{I}_I}(\widehat{C}_{i,J}) + 2 \sum_{I-J+1 \le i < k \le I} E\left[\left(C_{i,J} - \widehat{C}_{i,J}\right)\left(C_{k,J} - \widehat{C}_{k,J}\right) \middle| \mathcal{I}_I\right].$$

As above using the $\mathcal{I}_I$ measurability of $\hat{\beta}_j$ and $\hat{\mu}_i$ we obtain for $i < k$

$$E\left[\left(C_{i,J} - \widehat{C}_{i,J}\right)\left(C_{k,J} - \widehat{C}_{k,J}\right)\middle| \mathcal{I}_I\right] = \operatorname{Cov}\left(\sum_{j=I-i+1}^{J} X_{i,j}, \sum_{j=I-k+1}^{J} X_{k,j}\right)$$
$$+ \left(\hat{\mu}_i(1-\hat{\beta}_{I-i}) - \mu_i(1-\beta_{I-i})\right)\left(\hat{\mu}_k(1-\hat{\beta}_{I-k}) - \mu_k(1-\beta_{I-k})\right).$$

With the independence of the incremental claims $X_{i,j}$ we see that the covariance term is equal to zero and given $\mathcal{I}_I$ the last expression is an unknown constant. In order to estimate this unknown term we proceed as above and define

$$EE_{i,k} = \left(\hat{\mu}_i\left(1-\hat{\beta}_{I-i}\right) - \mu_i\left(1-\beta_{I-i}\right)\right)\left(\hat{\mu}_k\left(1-\hat{\beta}_{I-k}\right) - \mu_k\left(1-\beta_{I-k}\right)\right)$$
$$= \left((1-\hat{\beta}_{I-i})(\hat{\mu}_i - \mu_i) - \mu_i(\hat{\beta}_{I-i} - \beta_{I-i})\right)\left((1-\hat{\beta}_{I-k})(\hat{\mu}_k - \mu_k) - \mu_k(\hat{\beta}_{I-k} - \beta_{I-k})\right),$$

which is estimated by

$$\widehat{\mathrm{EE}}_{i,k} = (1-\hat{\beta}_{I-i})(1-\hat{\beta}_{I-k})\,\widehat{\operatorname{Cov}}(\hat{\mu}_i, \hat{\mu}_k) + \hat{\mu}_i\hat{\mu}_k\,\widehat{\operatorname{Cov}}(\hat{\beta}_{I-i}, \hat{\beta}_{I-k}),$$

where

$$\widehat{\operatorname{Cov}}(\hat{\mu}_i, \hat{\mu}_k) = \widehat{\operatorname{Corr}}(\hat{\mu}_i, \hat{\mu}_k)\sqrt{\widehat{\operatorname{Var}}(\hat{\mu}_i)\widehat{\operatorname{Var}}(\hat{\mu}_k)}, \tag{4.15}$$

and

$$\widehat{\operatorname{Cov}}\left(\hat{\beta}_{I-i}, \hat{\beta}_{I-k}\right) = \sum_{j=0}^{I-i}\sum_{l=0}^{I-k}\widehat{\operatorname{Cov}}(\hat{\gamma}_j, \hat{\gamma}_l),\ I-J+1 \le i,\ k \le I, \tag{4.16}$$

(see (4.14) and (4.9)).

Putting everything together we obtain the following estimate for the conditional MSEP:

Estimate 4.9. (MSEP aggregated accident years)
Under Model Assumptions 4.1 the conditional MSEP for aggregated accident years is estimated by

$$\widehat{\mathrm{msep}}_{\sum_{i=I-J+1}^{I} C_{i,J}|\mathcal{I}_I}\left(\sum_{i=I-J+1}^{I}\widehat{C}_{i,J}\right) = \sum_{i=I-J+1}^{I}\widehat{\mathrm{msep}}_{C_{i,J}|\mathcal{I}_I}\left(\widehat{C}_{i,J}\right)$$
$$+2\sum_{I-J+1\le i<k\le I}\left((1-\hat{\beta}_{I-i})(1-\hat{\beta}_{I-k})\,\widehat{\operatorname{Cov}}(\hat{\mu}_i, \hat{\mu}_k) + \hat{\mu}_i\hat{\mu}_k\,\widehat{\operatorname{Cov}}(\hat{\beta}_{I-i}, \hat{\beta}_{I-k})\right).$$

Remark 4.10. In the literature one often calculates the conditional MSEP given $\mathcal{D}_I$ instead of $\mathcal{I}_I$. This corresponds to a different interpretation. If the $\hat{\mu}_i$

are considered as observations one should condition on $\mathcal{I}_I$ otherwise if the $\hat{\mu}_i$ are considered as a priori information one should condition on $\mathcal{D}_I$. That is, in the latter case we are in a Bayesian framework, where μ_i are random variables with an a priori distribution and $\hat{\mu}_i$ is the expected value of μ_i. In the first case $\hat{\mu}_i$ is treated as an 'expert observation'. In the case where we condition on $\mathcal{D}_I$ we obtain for single accident years

$$\mathrm{msep}_{C_{i,J}|\mathcal{D}_I}(\widehat{C}_{i,J}) = \underbrace{\textstyle\sum_{j=I-i+1}^{J} \mathrm{Var}(X_{i,j})}_{PV_i} + \underbrace{E\Big[(\hat{\mu}_i(1-\hat{\beta}_{I-i}) - \mu_i(1-\beta_{I-i}))^2 \,\Big|\, \mathcal{D}_I\Big]}_{EE_i}.$$

In order to estimate the last term we consider the mean over all possible $\mathcal{D}_I$ (we could also use this procedure when we condition on $\mathcal{I}_I$ and the resulting estimates differ slightly).
With the approximation $E[\hat{\mu}_i(1-\hat{\beta}_{I-i})] \approx \mu_i(1-\beta_{I-i})$ we obtain

$$\begin{aligned} E\Big[\Big(\hat{\mu}_i(1-\hat{\beta}_{I-i}) - \mu_i(1-\beta_{I-i})\Big)^2\Big] &\approx \mathrm{Var}\Big(\hat{\mu}_i(1-\hat{\beta}_{I-i})\Big) \\ &= E\Big[\mathrm{Var}(\hat{\mu}_i(1-\hat{\beta}_{I-i})|\hat{\mu}_i)\Big] + \mathrm{Var}\Big(E\big[\hat{\mu}_i(1-\hat{\beta}_{I-i})\,\big|\,\hat{\mu}_i\big]\Big), \end{aligned}$$

and estimating the last term we obtain the following estimate for the estimation error

$$\widehat{\mathrm{EE}}_i = \Big(\hat{\mu}_i^2 + \widehat{\mathrm{Var}}(\hat{\mu}_i)\Big)\widehat{\mathrm{Var}}(1-\hat{\beta}_{I-i}) + (1-\hat{\beta}_{I-i})^2\,\widehat{\mathrm{Var}}(\hat{\mu}_i).$$

We therefore have the additional term $\widehat{\mathrm{Var}}(\hat{\mu}_i)\,\widehat{\mathrm{Var}}(1-\hat{\beta}_{I-i})$ in the estimation error. Similarly we have for aggregated accident years

$$EE_{i,k} = E\Big[\Big(\hat{\mu}_i(1-\hat{\beta}_{I-i}) - \mu_i(1-\beta_{I-i})\Big)\Big(\hat{\mu}_k(1-\hat{\beta}_{I-k}) - \mu_k(1-\beta_{I-k})\Big)\Big|\,\mathcal{D}_I\Big],$$

which is estimated by

$$\widehat{\mathrm{EE}}_{i,k} = \Big(1-\hat{\beta}_{I-i}\Big)\Big(1-\hat{\beta}_{I-k}\Big)\widehat{\mathrm{Cov}}(\hat{\mu}_i,\hat{\mu}_k) + \Big(\widehat{\mathrm{Cov}}(\hat{\mu}_i,\hat{\mu}_k) + \hat{\mu}_i\hat{\mu}_k\Big)\widehat{\mathrm{Cov}}(\hat{\beta}_{I-i},\hat{\beta}_{I-k}).$$

The resulting formulas for the conditional MSEP have the same form as the estimates derived in Mack [9] (but the estimates of the parameters and covariances are, of course, different).

4.2. General Over-dispersed Poisson Model

4.2.1. Model

In Model Assumptions 4.1 we have assumed that the dispersion parameter ϕ is constant over all development years. The assumption of a constant dispersion is quite restrictive and often not appropriate in practical applications. Therefore we consider in this subsection the general case, where the dispersion parameter varies between development years.

Model Assumptions 4.11. (General Over-dispersed Poisson Model)

GP1 Incremental claims $X_{i,j}$ are independent and there exist positive parameters $\phi_0, \ldots, \phi_J$, $\mu_0, \ldots, \mu_I$ and $\gamma_0, \ldots, \gamma_J$ with $\sum_{j=0}^{J} \gamma_j = 1$ such that $X_{i,j}/\phi_j \sim \text{Poisson}(\mu_i \gamma_j / \phi_j)$. In particular, we have

$$E[X_{i,j}] = \mu_i \gamma_j,$$
$$\text{Var}(X_{i,j}) = \phi_j \mu_i \gamma_j,$$

where ϕ_j is the dispersion parameter for development year j.

GP2 The a priori estimates $\hat{\mu}_i$ for $\mu_i = E[C_{i,J}]$ are unbiased and independent from $X_{l,j}$ for $0 \le l \le I$, $0 \le j \le J$.

4.2.2. Estimation of the Development Pattern

As in Section 4.1 we estimate γ_j with the MLE method using Lagrange multipliers. The Lagrange function is given by

$$\mathcal{L}_{\mathcal{D}_I}\left(\gamma_0, \ldots, \gamma_J, \kappa\right) = \sum_{i+j \le I} \frac{1}{\phi_j} \left(X_{i,j}\left(\log \mu_i + \log \gamma_j\right) - \mu_i \gamma_j\right) + \kappa\left(1 - \sum_{j=0}^{J} \gamma_j\right), \tag{4.17}$$

where κ is the Lagrange multiplier. For $0 \le j \le J$ the MLE $\tilde{\gamma}_j$ is the solution of

$$\frac{\partial \mathcal{L}_{\mathcal{D}_I}}{\partial \gamma_j} = \sum_{i=0}^{I-j} \frac{1}{\phi_j}\left(\frac{X_{i,j}}{\gamma_j} - \mu_i\right) - \kappa = 0, \tag{4.18}$$

and hence

$$\tilde{\gamma}_j = \frac{\sum_{i=0}^{I-j} \frac{1}{\phi_j} X_{i,j}}{\sum_{i=0}^{I-j} \frac{1}{\phi_j} \mu_i + \kappa} = \frac{X_{[I-j],j} / \phi_j}{\mu_{[I-j]} / \phi_j + \kappa}. \tag{4.19}$$

Remark 4.12. Note that (4.19) has the same structure as (4.1) with respect to the normalized random variables $Z_{i,j} = X_{i,j} / \phi_j$.

With the side constraint we obtain the following implicit equation for κ

$$\frac{\partial \mathcal{L}_{\mathcal{D}_I}}{\partial \kappa} = 1 - \sum_{j=0}^{J} \frac{X_{[I-j],j} / \phi_j}{\mu_{[I-j]} / \phi_j + \kappa} = 0.$$

It remains to estimate μ_i and ϕ_j.

For μ_i we use the a priori estimates $\hat{\mu}_i$. In order to estimate the dispersion parameters ϕ_j we consider $Y_{i,j} = X_{i,j}/\mu_i$ with

$$E[Y_{i,j}] = \gamma_j \text{ and } \operatorname{Var}(Y_{i,j}) = \phi_j \frac{\gamma_j}{\mu_i} = \frac{\sigma_j^2}{\mu_i},$$

where $\sigma_j^2 = \phi_j \gamma_j$. An unbiased estimate for σ_j^2 is given by

$$\tilde{\sigma}_j^2 = \frac{1}{I-j} \sum_{i=0}^{I-j} \mu_i (Y_{i,j} - \bar{Y}_j)^2, \; 0 \le j \le J, \; j \ne I, \tag{4.20}$$

with $\bar{Y}_j = \sum_{i=0}^{I-j} \frac{\mu_i}{\mu_{[I-j]}} Y_{i,j}$. If $J = I$ we use an extrapolation to obtain an estimate for σ_J^2, see Mack [8]. The dispersion parameter ϕ_j can then be estimated by

$$\hat{\phi}_j = \frac{\hat{\sigma}_j^2}{\hat{\gamma}_j}, \tag{4.21}$$

where $\hat{\sigma}_j^2$ is obtained from $\tilde{\sigma}_j^2$ by replacing the unknown μ_i's by the a priori estimates $\hat{\mu}_i$ and where

$$\hat{\gamma}_j = \frac{X_{[I-j],j} / \hat{\phi}_j}{\hat{\mu}_{[I-j]} / \hat{\phi}_j + \hat{\kappa}}, \tag{4.22}$$

with $\hat{\kappa} \in \left(-\min_{0 \le j \le J} \left\{ \frac{\hat{\mu}_{[I-j]}}{\hat{\phi}_j} \right\}, \infty \right)$ given by the implicit equation

$$1 - \sum_{j=0}^{J} \frac{X_{[I-j],j} / \hat{\phi}_j}{\hat{\mu}_{[I-j]} / \hat{\phi}_j + \hat{\kappa}} = 0.$$

Note that (4.21) is an implicit equation since $\hat{\phi}_j$ also appears in the equation for $\hat{\gamma}_j$. If we plug in $\hat{\phi}_j = \hat{\sigma}_j^2 / \hat{\gamma}_j$ in (4.22) we obtain

$$\hat{\gamma}_j = \frac{\hat{\gamma}_j X_{[I-j],j} / \hat{\sigma}_j^2}{\hat{\gamma}_j \hat{\mu}_{[I-j]} / \hat{\sigma}_j^2 + \hat{\kappa}},$$

and solving for $\hat{\gamma}_j$ we get

$$\hat{\gamma}_j = \frac{X_{[I-j],j}}{\hat{\mu}_{[I-j]}} - \hat{\kappa} \frac{\hat{\sigma}_j^2}{\hat{\mu}_{[I-j]}}.$$

With the side constraint $\sum_{j=0}^{J} \hat{\gamma}_j = 1$ we arrive at the following explicit solution

$$\hat{\gamma}_j = \frac{X_{[I-j],j}}{\hat{\mu}_{[I-j]}} + \frac{\hat{\sigma}_j^2 / \hat{\mu}_{[I-j]}}{\sum_{l=0}^{J} \hat{\sigma}_l^2 / \hat{\mu}_{[I-l]}} \left(1 - \sum_{l=0}^{J} \frac{X_{[I-l],l}}{\hat{\mu}_{[I-l]}} \right). \tag{4.23}$$

The cumulative development pattern is estimated by $\hat{\beta}_j = \sum_{k=0}^{j} \hat{\gamma}_k$, $0 \le j \le J$.

Remarks 4.13.

- For the case of a constant dispersion parameter ϕ we cannot use the above estimation procedure for ϕ. In that case $\sigma_j^2 / \gamma_j \equiv \phi$ but $\hat{\sigma}_j^2 / \hat{\gamma}_j$ might not be constant over all j. Moreover, if ϕ is constant over all development years, it does not make sense to consider the development years j separately for the estimation of ϕ.
- The estimators (4.23) do not fulfill equations (3.5) anymore.
- Contrary to the ODP Model, the MLEs based only on the observed data $\mathcal{D}_I$ and disregarding the μ_i are no longer identical to the CL forecasts. Instead, the MLEs $\hat{\mu}_i^{MLE}$ and the modified 'development pattern' $\hat{\beta}_j^{mod} = \sum_{k=0}^{j} \hat{\gamma}_k / \phi_k$ are proportional to the CL forecasts and the CL development pattern obtained from the modified data $X_{i,j} / \phi_j$.
- Denote again by

$$\hat{\gamma}_j^{(0)} = \frac{X_{[I-j],j}}{\hat{\mu}_{[I-j]}}$$

  the raw estimates. Then we again see how the smoothing from these raw estimates to the adjusted estimates $\hat{\gamma}_j$ looks like. Here the smoothing corrections are additive.
- A big deviation of $\sum_{j=0}^{J} \hat{\gamma}_j^{(0)}$ from 1 could be an indicator that the a priori estimates might be biased.
- A tail development $\hat{\gamma}_{J+1}$ can be incorporated in the same way as described in Remarks 4.6.

4.2.3. Covariance of the Estimated Development Pattern

As in Section 4.1 the covariance matrix of the $\hat{\gamma}_j$'s can be estimated using the Fisher information matrix. The log-likelihood function $l_{\mathcal{D}_I}$ is given by

$$\begin{aligned} l_{\mathcal{D}_I}(\gamma_0, \ldots, \gamma_{J-1}) &= \sum_{\substack{i+j \le I \\ j \le J-1}} \frac{1}{\phi_j} (X_{i,j} (\log \mu_i + \log \gamma_j) - \mu_i \gamma_j) \\ &+ \sum_{i=0}^{I-J} \frac{1}{\phi_J} \left(X_{i,J} \left(\log \mu_i + \log \left(1 - \sum_{j=0}^{J-1} \gamma_j \right) \right) - \mu_i \left(1 - \sum_{j=0}^{J-1} \gamma_j \right) \right) + r, \end{aligned} \tag{4.24}$$

where r contains all remaining terms, which do not depend on $\gamma = (\gamma_0, \ldots, \gamma_{J-1})$. Analogously to (4.6) we obtain for $0 \le j \le J-1$

$$\begin{aligned} H(\gamma)_{j,j} &= E\left[\left(\frac{\partial l_{\mathcal{D}_I}}{\partial \gamma_j} \right)^2 \right] = E\left[\left(\sum_{i=0}^{I-j} \frac{1}{\phi_j} \left(\frac{X_{i,j}}{\gamma_j} - \mu_i \right) - \sum_{i=0}^{I-J} \frac{1}{\phi_J} \left(\frac{X_{i,J}}{\gamma_J} - \mu_i \right) \right)^2 \right] \\ &= \frac{\mu_{[I-j]}}{\phi_j \gamma_j} + \frac{\mu_{[I-J]}}{\phi_J \gamma_J}, \end{aligned}$$

and for $0 \le j < k \le J - 1$ we get analogously to (4.7)

$$H(\gamma)_{k,j} = H(\gamma)_{j,k} = E\left[\frac{\partial l_{\mathcal{D}_I}}{\partial \gamma_j}\frac{\partial l_{\mathcal{D}_I}}{\partial \gamma_k}\right] = \frac{\mu_{[I-J]}}{\phi_J \gamma_J}.$$

The inversion of the Fisher information matrix is provided in Appendix A and we obtain for $0 \le j, k \le J - 1$ the approximation

$$\mathrm{Cov}(\hat{\gamma}_j, \hat{\gamma}_k) \approx (H(\gamma)^{-1})_{j,k} = \frac{\phi_j \gamma_j}{\mu_{[I-j]}}\left(1_{\{j=k\}} - \frac{\phi_k \gamma_k / \mu_{[I-k]}}{\sum_{l=0}^{J} \phi_l \gamma_l / \mu_{[I-l]}}\right). \quad (4.25)$$

By linearity we obtain analogous formulas for $\hat{\gamma}_J = 1 - \sum_{j=0}^{J-1} \hat{\gamma}_j$. Replacing the unknown parameters $\mu_{[I-j]}$, γ_j and ϕ_j by their estimates $\hat{\mu}_{[I-j]}$ (see Model Assumptions 4.11), $\hat{\gamma}_j$ (see (4.23)) and $\hat{\phi}_j$ (see (4.21)) we obtain the following estimates for the covariances

$$\widehat{\mathrm{Cov}}(\hat{\gamma}_j, \hat{\gamma}_k) = \frac{\hat{\phi}_j \hat{\gamma}_j}{\hat{\mu}_{[I-j]}}\left(1_{\{j=k\}} - \frac{\hat{\phi}_k \hat{\gamma}_k / \hat{\mu}_{[I-k]}}{\sum_{l=0}^{J} \hat{\phi}_l \hat{\gamma}_l / \hat{\mu}_{[I-l]}}\right), \; 0 \le j, k \le J. \quad (4.26)$$

For $\hat{\beta}_{I-i}$ we get

$$\widehat{\mathrm{Cov}}(\hat{\beta}_{I-i}, \hat{\beta}_{I-k}) = \sum_{j=0}^{I-i} \sum_{l=0}^{I-k} \widehat{\mathrm{Cov}}(\hat{\gamma}_j, \hat{\gamma}_l), \; I - J + 1 \le i, \, k \le I. \quad (4.27)$$

In the next section we give estimates for the conditional MSEP of the BF predictors $\widehat{C}_{i,J} = C_{i,I-i} + \hat{\mu}_i(1 - \hat{\beta}_{I-i})$ and $\sum_{i=I-J+1}^{I} \widehat{C}_{i,J}$ under Model Assumptions 4.11.

4.2.4. Conditional MSEP

The derivation of an estimate of the conditional MSEP for the BF predictor $\widehat{C}_{i,J}$ in the General ODP Model is analogous to Section 4.1.4. We obtain the following estimate for single accident years:

Estimate 4.14. (MSEP single accident year)

Under Model Assumptions 4.11 the conditional MSEP of the BF predictor $\widehat{C}_{i,J}$ is estimated by

$$\widehat{\mathrm{msep}}_{C_{i,J}|\mathcal{I}_I}(\widehat{C}_{i,J}) = \sum_{j=I-i+1}^{J} \hat{\phi}_j \hat{\mu}_i \hat{\gamma}_j + \widehat{\mathrm{Var}}(\hat{\mu}_i)(1 - \hat{\beta}_{I-i})^2 + \hat{\mu}_i^2 \, \widehat{\mathrm{Var}}(\hat{\beta}_{I-i}),$$

where $\widehat{\mathrm{Var}}(\hat{\beta}_{I-i})$ *is given in* (4.27) *and* $\widehat{\mathrm{Var}}(\hat{\mu}_i)$ *is given in* (4.14).

Using the same ideas as in the derivation of Estimate 4.9 we obtain the following estimate for aggregated accident years:

Estimate 4.15. (MSEP aggregated accident years)

Under Model Assumptions 4.11 the conditional MSEP for aggregated accident years is estimated by

$$\widehat{\text{msep}}_{\sum_{i=I-J+1}^{I} C_{i,J}|\mathcal{I}_I}\left(\sum_{i=I-J+1}^{I} \widehat{C}_{i,J}\right) = \sum_{i=I-J+1}^{I} \widehat{\text{msep}}_{C_{i,J}|\mathcal{I}_I}\left(\widehat{C}_{i,J}\right)$$
$$+2 \sum_{I-J+1 \le i < k \le I} \left((1-\hat{\beta}_{I-i})(1-\hat{\beta}_{I-k})\, \widehat{\text{Cov}}(\hat{\mu}_i, \hat{\mu}_k) + \hat{\mu}_i \hat{\mu}_k\, \widehat{\text{Cov}}(\hat{\beta}_{I-i}, \hat{\beta}_{I-k})\right),$$

where $\widehat{\text{Cov}}(\hat{\beta}_{I-i}, \hat{\beta}_{I-k})$ *is given in* (4.27) *and* $\widehat{\text{Cov}}(\hat{\mu}_i, \hat{\mu}_k)$ *is given in* (4.15).

4.3. Normal Model

4.3.1. Model

As mentioned in Remarks 4.2 the ODP Model can only be used if the incremental claims $X_{i,j}$ are positive, which is often appropriate for claims payments. In contrast, incurred losses increments are sometimes negative and therefore an ODP model is not suitable in this case. The following model does not have these restrictions.

Model Assumptions 4.16 (Normal Model)

N1 Incremental claims $X_{i,j}$ are independent and normally distributed and there exist parameters $\mu_0, \ldots, \mu_I, \sigma_0^2, \ldots, \sigma_J^2$ and $\gamma_0, \ldots, \gamma_J$ with $\sum_{j=0}^{J} \gamma_j = 1$ such that

$$E[X_{i,j}] = \mu_i \gamma_j,$$
$$\text{Var}(X_{i,j}) = \mu_i \sigma_j^2,$$

where σ_j^2 is strictly positive.

N2 The a priori estimates $\hat{\mu}_i$ of $\mu_i = E[C_{i,J}]$ are unbiased and independent from $X_{l,j}$ for $0 \le l \le I$, $0 \le j \le J$.

In the following we derive the MLEs and the corresponding conditional MSEP under Model Assumptions 4.16.

4.3.2. Estimation of the Development Pattern

As in the ODP Model we first assume that the μ_i are known and calculate the MLEs for the γ_j using Lagrange multipliers. In the Normal Model this procedure allows us to find explicit solutions for the estimates. The Lagrange function is given by

$$\mathcal{L}_{\mathcal{D}_I}(\gamma_0, \ldots, \gamma_J, \kappa) = \sum_{i+j \le I} - \frac{(X_{i,j} - \mu_i \gamma_j)^2}{2\mu_i \sigma_j^2} + \kappa\left(1 - \sum_{j=0}^{J} \gamma_j\right),$$

where κ is the Lagrange multiplier. For $0 \le j \le J$ we have

$$\frac{\partial \mathcal{L}_{\mathcal{D}_I}}{\partial \gamma_j} = \sum_{i=0}^{I-j} \frac{X_{i,j} - \mu_i \gamma_j}{\sigma_j^2} - \kappa = 0,$$

and therefore

$$\tilde{\gamma}_j = \frac{X_{[I-j],j} - \kappa \sigma_j^2}{\mu_{[I-j]}}. \tag{4.28}$$

With the side constraint

$$1 = \sum_{j=0}^{J} \tilde{\gamma}_j = \sum_{j=0}^{J} \frac{X_{[I-j],j}}{\mu_{[I-j]}} - \sum_{j=0}^{J} \kappa \frac{\sigma_j^2}{\mu_{[I-j]}},$$

it follows that

$$\kappa = \frac{\sum_{j=0}^{J} \frac{X_{[I-j],j}}{\mu_{[I-j]}} - 1}{\sum_{j=0}^{J} \frac{\sigma_j^2}{\mu_{[I-j]}}}.$$

We insert κ in equation (4.28) and obtain the MLE for γ_j

$$\tilde{\gamma}_j = \frac{X_{[I-j],j}}{\mu_{[I-j]}} + \frac{\sigma_j^2 / \mu_{[I-j]}}{\sum_{l=0}^{J} \sigma_l^2 / \mu_{[I-l]}} \left(1 - \sum_{l=0}^{J} \frac{X_{[I-l],l}}{\mu_{[I-l]}}\right). \tag{4.29}$$

The μ_i and σ_j^2 in formula (4.29) are unknown and have to be replaced by estimators. For μ_i we insert $\hat{\mu}_i$ and σ_j^2 can be estimated analogously as in the General ODP Model, that is, by replacing the μ_i in formula (4.20) by $\hat{\mu}_i$.

Remark 4.17. Using MLEs for σ_j^2 results in a system of equations for γ_j and σ_j that is only iteratively solvable.

We arrive at the final estimate

$$\hat{\gamma}_j = \frac{X_{[I-j],j}}{\mu_{[I-j]}} + \frac{\hat{\sigma}_j^2 / \hat{\mu}_{[I-j]}}{\sum_{l=0}^{J} \hat{\sigma}_l^2 / \hat{\mu}_{[I-l]}} \left(1 - \sum_{l=0}^{J} \frac{X_{[I-l],l}}{\mu_{[I-l]}}\right), \tag{4.30}$$

which is exactly the same estimator as in the General ODP Model (see (4.23)). The cumulative development pattern is estimated by $\hat{\beta}_j = \sum_{k=0}^{j} \hat{\gamma}_k$, $0 \le j \le J$.

Remark 4.18. A tail development $\hat{\gamma}_{J+1}$ can be incorporated as described in Remarks 4.6.

4.3.3. Covariance Matrix of the Estimated Development Pattern

Since the estimated development pattern is the same as in the General ODP Model we could use the estimated covariance matrix of the $\hat{\gamma}_j$ derived in Section 4.2.3. However, the $\hat{\gamma}_j$ are estimated with the MLE method and for consistency with the derivations in the other models we use the Fisher information matrix to estimate the covariance matrix of the $\hat{\gamma}_j$. The likelihood function for the Normal Model is given by

$$l_{\mathcal{D}_I}(\gamma_0, \dots, \gamma_{J-1}) = \sum_{\substack{i+j \le I \\ j \le J-1}} -\frac{(X_{i,j} - \mu_i \gamma_j)^2}{2\mu_i \sigma_j^2} + \sum_{i=0}^{I-J} -\frac{\left(X_{i,J} - \mu_i \left(1 - \sum_{j=0}^{J-1} \gamma_j\right)\right)^2}{2\mu_i \sigma_J^2} + r,$$

where r contains all remaining terms, which do not depend on $\gamma = (\gamma_0, \dots, \gamma_{J-1})$. For the entries of the Fisher information matrix we get analogously to Section 4.1.2.

$$H(\gamma)_{j,k} = \frac{\mu_{[I-j]}}{\sigma_j^2} \left(\mathbb{1}_{\{j=k\}} + \frac{\mu_{[I-J]}/\sigma_J^2}{\mu_{[I-j]}/\sigma_j^2} \right),$$

where $0 \le j, k \le J-1$. With the calculation provided in Appendix A we obtain the entries of the inverse of the Fisher information matrix and for $0 \le j$, $k \le J-1$ we get the approximations

$$\mathrm{Cov}(\hat{\gamma}_j, \hat{\gamma}_k) \approx \left(H(\gamma)^{-1}\right)_{j,k} = \frac{\sigma_j^2}{\mu_{[I-j]}} \left(\mathbb{1}_{\{j=k\}} - \frac{\sigma_k^2/\mu_{[I-k]}}{\sum_{l=0}^{J} \sigma_l^2/\mu_{[I-l]}} \right). \tag{4.31}$$

By linearity we obtain the same formulas for the variance and covariances of $\hat{\gamma}_J$. By replacing the unknown parameters $\mu_{[I-j]}$ and σ_j^2 by their estimates we obtain exactly the same formulas as in the General ODP Model, that is,

$$\widehat{\mathrm{Cov}}(\hat{\gamma}_j, \hat{\gamma}_k) = \frac{\hat{\sigma}_j^2}{\hat{\mu}_{[I-j]}} \left(\mathbb{1}_{\{j=k\}} - \frac{\hat{\sigma}_k^2/\hat{\mu}_{[I-k]}}{\sum_{l=0}^{J} \hat{\sigma}_l^2/\hat{\mu}_{[I-l]}} \right), \quad 0 \le j, k \le J. \tag{4.32}$$

The covariances of the $\hat{\beta}_{I-i}$ are estimated by

$$\widehat{\mathrm{Cov}}\left(\hat{\beta}_{I-i}, \hat{\beta}_{I-k}\right) = \sum_{j=0}^{I-i} \sum_{l=0}^{I-k} \widehat{\mathrm{Cov}}(\hat{\gamma}_j, \hat{\gamma}_l), \; I-J+1 \leq i, \; k \leq I. \tag{4.33}$$

Remark 4.19. The coincidence of the estimated covariance matrices in the General ODP Model and the Normal Model is of course meaningful, however not obvious because of the approximations used.

In the next section we give an estimate for the conditional MSEP of the BF predictor in the Normal Model.

4.3.4. Conditional MSEP

Since the estimates in the Normal Model coincide with the estimates in the General ODP Model we also obtain the same estimates for the conditional MSEP:

Estimate 4.20. (MSEP single accident year)

Under Model Assumptions 4.16 the conditional MSEP of the BF predictor $\widehat{C}_{i,J}$ is estimated by

$$\widehat{\mathrm{msep}}_{C_{i,J}|\mathcal{I}_I}(\widehat{C}_{i,J}) = \sum_{j=I-i+1}^{J} \hat{\mu}_i \hat{\sigma}_j^2 + \widehat{\mathrm{Var}}(\hat{\mu}_i)(1-\hat{\beta}_{I-i})^2 + \hat{\mu}_i^2 \, \widehat{\mathrm{Var}}(\hat{\beta}_{I-i}),$$

where $\widehat{\mathrm{Var}}(\hat{\beta}_{I-i})$ *is given in* (4.33) *and* $\widehat{\mathrm{Var}}(\hat{\mu}_i)$ *is given in* (4.14).

For aggregated accident years we obtain in the Normal Model:

Estimate 4.21. (MSEP aggregated accident years)

Under Model Assumptions 4.16 the conditional MSEP for aggregated accident years is estimated by

$$\widehat{\mathrm{msep}}_{\sum_{i=I-J+1}^{I} C_{i,J}|\mathcal{I}_I}\left(\sum_{i=I-J+1}^{I} \widehat{C}_{i,J}\right) = \sum_{i=I-J+1}^{I} \widehat{\mathrm{msep}}_{C_{i,J}|\mathcal{I}_I}(\widehat{C}_{i,J})$$
$$+ 2 \sum_{I-J+1 \leq i < k \leq I} \left((1-\hat{\beta}_{I-i})(1-\hat{\beta}_{I-k}) \, \widehat{\mathrm{Cov}}(\hat{\mu}_i, \hat{\mu}_k) + \hat{\mu}_i \hat{\mu}_k \, \widehat{\mathrm{Cov}}\left(\hat{\beta}_{I-i}, \hat{\beta}_{I-k}\right) \right),$$

where $\widehat{\mathrm{Cov}}(\hat{\beta}_{I-i}, \hat{\beta}_{I-k})$ *is given in* (4.33) *and* $\widehat{\mathrm{Cov}}(\hat{\mu}_i, \hat{\mu}_k)$ *is given in* (4.15).

Remark 4.22. Analogous estimators can be derived in the more general Tweedie's exponential dispersion family models, similar to Alai-Wüthrich [1].

5. Conclusions and Remarks

In this paper we have investigated the question how the development pattern should be estimated in the BF method and we have derived the corresponding MSEP of the ultimate claim prediction. For this purpose we have considered three distributional models. For these models we have been able to find estimators for the development pattern γ_j which are consistent with the BF philosophy. Moreover, we have found formulas for smoothing from the raw estimates $\hat{\gamma}_j^{(0)}$ to the final estimates $\hat{\gamma}_j$ and we have been able to find explicit formulas for the correlation matrix of these estimates in terms of the inverse Fisher information matrix.

The ODP Model is presumably not an adequate model for most practical cases. The General ODP Model is a reasonable model for claims payments modelling and the Normal Model can be used for incurred claims studies. For the latter two models we have found the same estimators. But also in the case where the distributional assumptions are not fully satisfied, we suggest applying these estimators because currently there are no estimators available from which we know that they perform better. Therefore we suggest

– to estimate the development pattern by means of formula (4.30)
– to estimate the correlation matrix of these estimates by means of formula (4.32)
– to estimate the conditional MSEP of the ultimate claim by means of Estimate 4.20 (single accident year) and Estimate 4.21 (aggregated accident years).

6. Numerical Examples

The data for the numerical examples are from a Swiss insurance company and for confidentiality purposes the figures are scaled with a constant. The a priori estimates are obtained from pricing. To be more precise $\hat{\mu}_i$ corresponds to the initial forecast of the expected ultimate claim at the end of year $i-1$ and there is not done any repricing afterwards. We consider claims payments data from industrial property insurance and incurred losses data from motor liability insurance. Industrial property insurance is a short tailed line of business meaning that the development is usually finished after short time. On the contrary, motor liability is a long tailed line of business, that is, we have longer settlement periods.

For comparison we also give the results obtained with the CL method. The conditional MSEP for the CL method is calculated according to the distribution-free model by Mack [8].

6.1. Industrial Property Insurance

Let us first consider data from industrial property, which is given in Table 5 in Appendix B. We apply the estimators suggested in Section 5 and refer to them as BF in the following tables. In this example we additionally give the results obtained in the ODP Model (BF ODP) and the results obtained using Mack's model [9] (BF Mack). With the trapezoid of claims payments we obtain estimated coefficients of variation of the a priori estimates of $\widehat{\text{CoVa}}(\hat{\mu}_i) = 5.25\%$ in the ODP Model and $\widehat{\text{CoVa}}(\hat{\mu}_i) = 4.56\%$ for BF. The resulting development patterns and the reserves are given in Table 1. The development patterns are obtained from formulas (4.3) and (4.30) without manual smoothing. Note that all three BF development patterns are close to the CL development pattern. Moreover, we calculated the 'usual' BF reserves, which are obtained with the BF method using the CL development pattern. Due to the rather small value on the diagonal in the newest accident year the BF reserves are higher than

TABLE 1

INDUSTRIAL PROPERTY, CLAIMS PAYMENTS:
ESTIMATED DEVELOPMENT PATTERN AND RESERVES.

	Dev. pattern $\hat{\beta}_{I-i}$				Estimated reserves				
i	BF ODP	BF	BF Mack	CL	BF ODP	BF	BF Mack	CL	'usual' BF
9	99.77%	99.78%	99.77%	99.78%	268	257	261	230	246
10	99.55%	99.57%	99.56%	99.59%	505	481	492	290	467
11	99.25%	99.29%	99.27%	99.29%	766	731	751	636	725
12	98.45%	98.48%	98.47%	98.50%	1'501	1'468	1'479	1'313	1'454
13	94.08%	94.24%	94.13%	94.14%	5'830	5'677	5'786	5'946	5'774
14	60.21%	60.59%	60.31%	60.40%	38'611	38'240	38'520	34'502	38'426
Total					**47'481**	**46'854**	**47'288**	**42'916**	**47'091**

TABLE 2

INDUSTRIAL PROPERTY, CLAIMS PAYMENTS:
ESTIMATED CONDITIONAL MSEP$^{1/2}$ AND COEFFICIENT OF VARIATION.

	msep$^{1/2}$				CoVa			
i	BF ODP	BF	BF Mack	CL	BF ODP	BF	BF Mack	CL
9	410	373	373	341	152.6%	145.1%	143.3%	148.3%
10	560	435	435	325	110.9%	90.3%	88.4%	112.1%
11	685	508	509	457	89.3%	69.5%	67.8%	72.0%
12	953	1'097	1'099	1'064	63.5%	74.7%	74.3%	81.0%
13	1'886	1'861	1'876	1'946	32.3%	32.8%	32.4%	32.7%
14	5'133	6'257	6'516	6'073	13.3%	16.4%	16.9%	17.6%
Total	**5'875**	**6'829**	**6'988**	**6'587**	**12.4%**	**14.6%**	**14.8%**	**15.3%**

TABLE 3

INDUSTRIAL PROPERTY, CLAIMS PAYMENTS:
ESTIMATED PROCESS STANDARD DEVIATION AND SQUARE ROOT OF THE ESTIMATION ERROR.

	Process std. dev.				(Estimation error)$^{1/2}$			
i	BF ODP	BF	BF Mack	CL	BF ODP	BF	BF Mack	CL
9	385	351	351	323	139	126	127	111
10	529	410	410	313	185	146	146	86
11	651	483	483	438	211	160	160	133
12	911	1'053	1'053	1'024	279	310	312	286
13	1'796	1'777	1'788	1'869	575	554	567	542
14	4'622	5'874	6'067	5'885	2'232	2'156	2'377	1'501
Total	**5'126**	**6'268**	**6'453**	**6'291**	**2'871**	**2'710**	**2'682**	**1'952**

the CL reserves. In Table 2 the conditional MSEP and the corresponding coefficient of variation are given. In BF Mack we applied the coefficient of variation $\widehat{\text{CoVa}}(\hat{\mu}_i) = 4.56\%$ from the Normal Model. The process standard deviation and estimation error are given in Table 3. Note that the difference between the conditional MSEP in the ODP Model and in BF in the newest accident year comes from the process variance. More precisely, it is due to the rather small value $\hat{\phi}\hat{\gamma}_1 = 187$ compared to $\hat{\sigma}_1^2 = 323$ appearing in the corresponding process variances. Similarly we have $\hat{s}_1^2 = 347$ for the corresponding term in Mack's model. The assumption of a constant dispersion parameter seems therefore questionable.

The results of BF are very close to the results of BF Mack, especially for older accident years. For newer accident years the process variance term contains more parameter estimates $\hat{\sigma}_j^2$ and $\hat{s}_j^2$, respectively. The differences in the estimation error are mainly due to the different estimation of the variances $\widehat{\text{Var}}(\hat{\beta}_{I-i})$. The additional term $\widehat{\text{Var}}(\hat{\mu}_i)\widehat{\text{Var}}(\hat{\beta}_{I-i})$ in Mack's formula for the estimation error is negligible compared to the other terms.

6.2. Motor Liability Insurance

The data from motor liability are given in Table 6 and Table 7 in Appendix B. Note that the observations $C_{0,0}$, $C_{1,0}$ and $C_{0,1}$ are missing. But Table 7 is not really an incomplete triangle because $C_{0,j}$, $j \geq 2$ and $C_{1,j}$, $j \geq 1$ contain all claims of accident years 0 and 1, respectively, that is, also the closed one's. The results obtained with the data from Table 7 are given in Table 4. We apply the estimators suggested in Section 5 and obtain the estimate $\widehat{\text{CoVa}}(\hat{\mu}_i) = 9.47\%$. The reserves are calculated using formula (2.2) with the diagonal claims payments given in Table 6. The BF reserves are rather high compared to the CL

TABLE 4

MOTOR LIABILITY, INCURRED LOSSES:
ESTIMATED DEVELOPMENT PATTERN, RESERVES, CONDITIONAL MSEP$^{1/2}$ AND COEFFICIENT OF VARIATION.

	Dev. pattern $\hat{\beta}_{I-i}$		Estimated reserves			msep$^{1/2}$		CoVa	
i	BF	CL	BF	CL	'usual' BF	BF	CL	BF	CL
5	99.98%	100.01%	7'236	7'130	7'130	948	926	13.1%	13.0%
6	99.90%	99.93%	11'504	11'337	11'358	1'159	1'100	10.1%	9.7%
7	100.08%	100.12%	14'520	14'341	14'326	1'322	1'283	9.1%	8.9%
8	100.05%	100.12%	9'141	8'837	8'804	1'952	1'914	21.4%	21.7%
9	99.80%	99.93%	22'902	22'274	22'297	2'931	2'847	12.8%	12.8%
10	99.80%	99.94%	26'574	25'870	25'888	3'166	3'079	11.9%	11.9%
11	99.66%	99.83%	39'181	38'354	38'357	3'638	3'690	9.3%	9.6%
12	99.71%	99.90%	42'358	41'487	41'519	3'687	3'599	8.7%	8.7%
13	99.54%	99.78%	43'398	42'296	42'303	4'500	4'553	10.4%	10.8%
14	99.49%	99.79%	64'401	63'116	63'092	5'126	5'275	8.0%	8.4%
15	99.57%	99.93%	71'833	70'258	70'253	5'765	5'897	8.0%	8.4%
16	99.29%	99.69%	90'656	88'897	88'864	6'373	6'549	7.0%	7.4%
17	98.91%	99.44%	102'764	100'412	100'425	7'707	7'824	7.5%	7.8%
18	98.72%	99.41%	90'343	87'150	87'332	9'092	8'906	10.1%	10.2%
19	98.67%	99.53%	142'365	138'381	138'445	10'693	10'738	7.5%	7.8%
20	98.36%	99.60%	142'585	136'460	136'744	13'654	12'774	9.6%	9.4%
21	98.40%	100.04%	126'460	118'925	118'880	15'938	14'291	12.6%	12.0%
22	98.48%	100.48%	125'213	116'434	115'704	18'159	15'146	14.5%	13.0%
23	99.69%	102.21%	153'111	144'103	141'922	19'740	17'725	12.9%	12.3%
24	100.66%	101.41%	205'725	203'149	202'454	24'655	26'502	12.0%	13.0%
25	102.59%	101.49%	262'721	267'507	267'278	28'535	33'753	10.9%	12.6%
Total			**1'794'990**	**1'746'719**	**1'743'375**	**61'926**	**64'813**	**3.4%**	**3.7%**

and to the 'usual' BF reserves. Interestingly, the reserves obtained by the 'usual' BF method using the CL development pattern are close to the CL reserves and even slightly lower. This means that the reason for the higher reserves obtained with BF cannot be conservative a priori estimates $\hat{\mu}_i$, but is rather the new BF consistent estimate of the development pattern. This example shows, that the way how the development pattern is estimated in the BF method can have a big impact on the resulting reserves. As already pointed out previously, the usual way of simply using the CL development pattern is not consistent with the BF philosophy.

APPENDIX

A. Proofs

Inversion of the Fisher Information Matrix. In order to calculate the inverse of the Fisher information matrix in (4.8), (4.25) and (4.31) we need to invert a matrix of the form

$$A = \begin{pmatrix} a_0+1 & 1 & 1 & \cdots & 1 \\ 1 & a_1+1 & 1 & \cdots & 1 \\ 1 & 1 & a_2+1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & a_{J-1}+1 \end{pmatrix},$$

where $a_0, a_1, \ldots, a_{J-1} \in \mathbb{R} \setminus \{0\}$. We number the rows and columns of A starting from 0 and denote the entry in row j and column k by $A_{j,k}$, $0 \le j, k \le J-1$, that is, $A_{j,k} = 1 + a_j \mathbb{1}_{\{j=k\}}$. We claim that

$$(A^{-1})_{j,k} = \frac{1}{a_j}\left(\mathbb{1}_{\{j=k\}} - \frac{a_0}{d a_k}\right), \; 0 \le j, k \le J-1,$$

where $d = a_0\left(1 + \sum_{j=0}^{J-1} \frac{1}{a_j}\right)$.

Proof. For $0 \le j, k \le J-1$ we have

$$\begin{aligned}(A \cdot A^{-1})_{j,k} &= \sum_{l=0}^{J-1} (1 + a_j \cdot \mathbb{1}_{\{j=l\}}) \frac{1}{a_l}\left(\mathbb{1}_{\{l=k\}} - \frac{a_0}{d a_k}\right) \\ &= \frac{1}{a_k} - \sum_{l=0}^{J-1} \frac{a_0}{a_l d a_k} + \mathbb{1}_{\{j=k\}} - \frac{a_0}{d a_k} \\ &= \mathbb{1}_{\{j=k\}} + \frac{1}{a_k}\left(1 - \frac{a_0}{d}\left(\sum_{l=0}^{J-1} \frac{1}{a_l} + 1\right)\right) \\ &= \mathbb{1}_{\{j=k\}},\end{aligned}$$

where we used the definition of d in the last equation. □

Proof of Equation (4.13). With $\beta_j = 1$ for $j > J$ we have

$$\text{CoVa}^2(\hat{\Pi}_I) = \text{CoVa}^2\left(\sum_{i=0}^{I} \hat{\mu}_i \beta_{I-i}\right) = \frac{\text{Var}\left(\sum_{i=0}^{I} \hat{\mu}_i \beta_{I-i}\right)}{\left(\sum_{i=0}^{I} \mu_i \beta_{I-i}\right)^2}$$

$$= \frac{\sum_{i=0}^{I} \beta_{I-i}^2 \text{Var}(\hat{\mu}_i) + 2\sum_{i<k} \beta_{I-i}\beta_{I-k} \text{Corr}(\hat{\mu}_i, \hat{\mu}_k) \mu_i \text{CoVa}(\hat{\mu}_i) \mu_k \text{CoVa}(\hat{\mu}_k)}{\left(\sum_{i=0}^{I} \mu_i \beta_{I-i}\right)^2}$$

$$= \frac{c^2\left(\sum_{i=0}^{I}\beta_{I-i}^2\mu_i^2 + 2\sum_{i<k}\beta_{I-i}\beta_{I-k}\mu_i\mu_k \operatorname{Corr}(\hat{\mu}_i,\hat{\mu}_k)\right)}{\left(\sum_{i=0}^{I}\mu_i\beta_{I-i}\right)^2}$$

$$= c^2\,\frac{\left(\left(\sum_{i=0}^{I}\beta_{I-i}\mu_i\right)^2 - 2\sum_{i<k}\beta_{I-i}\beta_{I-k}\mu_i\mu_k\left(1-\operatorname{Corr}(\hat{\mu}_i,\hat{\mu}_k)\right)\right)}{\left(\sum_{i=0}^{I}\mu_i\beta_{I-i}\right)^2}.$$

□

Proof of Remark 4.7. Estimating the a priori loss ratios with a moving average of the ultimate loss ratios over n years and assuming that $v_i \equiv v$ and $\operatorname{Var}(U_i/v) = \sigma^2/v^2$ yields the following formulas

$$\hat{\mu}_i = v\hat{q}_i = \frac{\sum_{l=i}^{n+i-1} U_l}{n}, \quad 0 \le i \le I,$$

where U_i denotes the ultimate claim amount for accident year i. For $i \le k$ it follows with the independence of accident years

$$\operatorname{Cov}(\hat{\mu}_i,\hat{\mu}_k) = \frac{1}{n^2}\operatorname{Cov}\left(\sum_{l=i}^{n+i-1} U_l, \sum_{l=k}^{n+k-1} U_l\right) = \frac{1}{n^2}\sum_{l=k}^{n+i-1}\sigma^2 = \frac{n+i-k}{n^2}\sigma^2,$$

and since $\operatorname{Var}(\hat{\mu}_i) = \sigma^2/n$ we obtain

$$\operatorname{Corr}(\hat{\mu}_i,\hat{\mu}_k) = \frac{n-|i-k|}{n^2\sigma^2/n}\sigma^2 = \frac{n-|i-k|}{n}.$$

□

B. Data

TABLE 5

INDUSTRIAL PROPERTY, CUMULATIVE PAYMENTS AND A PRIORI ESTIMATES $\hat{\mu}_i$.

i/j	0	1	2	3	4	5	ultimate	$\hat{\mu}_i$
0	52’572	76’651	80’044	80’524	80’870	81’459	81’587	81’552
1	58’623	89’190	94’040	95’592	95’637	95’765	95’898	87’138
2	71’086	108’235	110’410	110’917	110’883	111’092	111’049	100’276
3	58’236	86’079	91’586	90’303	90’490	90’507	90’372	99’319
4	66’661	108’829	113’347	114’785	115’656	115’756	116’481	102’035
5	56’059	90’688	96’389	96’661	97’015	97’160	97’542	100’963
6	52’443	87’856	91’063	91’846	92’414	92’855	92’920	101’178
7	67’307	102’881	107’783	108’279	108’644	108’844	109’599	102’764
8	67’829	98’815	102’008	102’374	102’775	102’868	102’792	111’570
9	69’259	100’684	104’879	106’717	106’602	106’668		114’284
10	41’714	66’880	69’390	69’697	69’869			113’055
11	54’717	82’924	86’781	89’270				102’519
12	46’429	79’564	86’174					96’879
13	55’001	95’511						98’517
14	52’630							97’041

TABLE 6

MOTOR LIABILITY, DIAGONAL PAYMENTS.

i	$C_{i,I-i}^{paid}$
0	268'392
1	286'310
2	272'781
3	341'679
4	337'137
5	381'388
6	399'724
7	424'117
8	419'528
9	411'082
10	410'387
11	435'980
12	372'513
13	411'770
14	380'622
15	384'000
16	364'883
17	336'143
18	317'801
19	300'874
20	262'034
21	242'768
22	206'808
23	199'872
24	185'856
25	132'116

TABLE 7

MOTOR LIABILITY, INCURRED LOSSES AND A PRIORI ESTIMATES $\hat{\mu}_i$.

i/j	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	ultimate	$\hat{\mu}_i$
0			284'451	277'284	268'633	265'993	263'702	263'260	262'612	262'596	264'525	264'165	265'054	266'590	267'035	267'291	267'217	270'073	269'350	269'308	268'740	268'392	242'359
1		314'004	307'747	289'276	283'281	281'033	277'636	276'586	277'460	277'042	279'303	281'596	284'693	284'264	284'787	284'472	285'261	286'117	286'307	285'422	285'683	286'310	270'229
2	321'578	309'394	289'928	282'004	275'502	270'268	271'434	269'822	266'568	267'388	268'193	268'081	267'289	266'985	266'706	266'820	267'891	270'210	272'464	272'152	272'473	272'781	309'063
3	371'156	367'120	346'871	339'593	342'688	339'325	332'822	333'296	329'576	333'859	337'626	338'856	338'417	341'972	341'159	342'548	342'949	343'067	343'512	342'166	342'871	341'679	337'683
4	370'661	369'491	358'572	344'401	346'665	342'370	344'807	344'529	346'520	344'774	345'346	344'592	339'747	340'214	338'824	338'687	338'119	337'285	336'834	336'271	336'647	337'137	348'026
5	419'289	416'218	393'048	384'661	383'508	376'001	382'602	383'968	384'388	383'803	387'951	389'023	386'576	389'510	388'477	388'833	389'687	390'600	389'190	388'147	388'548		398'224
6	437'568	427'975	415'222	407'791	406'460	418'175	411'091	406'724	410'959	413'780	414'044	410'973	410'142	409'120	410'870	414'860	412'483	410'468	410'964	410'768			439'606
7	450'610	448'418	431'294	424'361	432'825	427'773	436'947	434'592	437'282	437'315	438'579	439'037	440'049	439'942	439'561	437'789	437'218	439'781	438'979				450'296
8	468'979	450'942	432'591	435'162	423'583	420'876	424'588	424'632	429'180	431'788	429'673	433'284	430'848	432'002	431'379	429'732	429'539	428'875					456'387
9	443'189	434'315	433'325	425'332	424'921	422'432	421'233	422'904	424'484	426'661	431'462	431'676	434'368	433'599	431'820	432'994	433'037						463'645
10	435'307	422'149	412'660	405'190	402'367	402'597	411'096	412'873	420'973	432'325	437'067	433'610	435'824	435'656	434'796	436'011							468'488
11	475'948	438'817	432'413	436'710	436'904	443'156	448'347	463'171	468'412	476'070	474'593	473'324	477'058	472'283	473'524								476'198
12	447'021	422'678	405'919	399'462	400'047	398'297	406'939	404'834	409'056	411'421	412'002	410'739	409'744	413'587									445'631
13	457'229	444'054	436'390	436'853	442'292	453'494	456'363	460'272	459'591	456'975	455'336	454'500	453'068										457'050
14	462'989	464'776	447'833	432'893	432'124	442'743	451'994	451'534	450'528	450'845	448'398	442'810											432'273
15	484'915	468'800	454'958	447'601	461'106	470'358	465'346	468'879	461'537	456'753	453'919												447'010
16	462'028	429'610	438'929	454'797	468'116	468'721	469'907	463'823	459'524	452'385													443'103
17	450'908	456'030	476'259	483'129	476'952	464'941	453'391	445'089	434'103														438'830
18	426'385	428'504	456'796	449'886	445'397	432'021	412'353	402'565															435'961
19	461'078	477'458	480'960	471'869	462'978	444'670	437'203																453'019
20	444'123	430'684	433'664	419'422	403'126	396'903																	469'713
21	433'830	407'931	393'723	371'800	361'853																		461'548
22	418'202	374'855	338'598	324'790																			475'731
23	426'853	373'282	351'590																				442'506
24	410'810	394'477																					438'441
25	405'597																						414'974

References

[1] Alai, D.H. and Wüthrich, M.V. (2009) Taylor approximations for model uncertainty within the Tweedie exponential dispersion family. *ASTIN Bulletin* **39(2)**, 453-478.

[2] Alai, D.H., Merz, M. and Wüthrich, M.V. (2010) Prediction uncertainty in the Bornhuetter-Ferguson claims reserving method: revisited. *Annals of Actuarial Science* **5(1)**, 7-17.

[3] Bornhuetter, R.L. and Ferguson, R.E. (1972) The actuary and IBNR. *Proc. CAS*, **LIX**, 181-195.

[4] Bühlmann, H. and Straub, E. (1983) Estimation of IBNR reserves by the methods chain ladder, Cape Cod and complementary loss ratio. International Summer School 1983, unpublished.

[5] England, P.D. and Verrall, R.J. (2002) Stochastic claims reserving in general insurance. *British Actuarial Journal* **8(3)**, 443-518.

[6] Hachemeister, C. and Stanard, J. (1975) IBNR claims count estimation with static lag functions. Presented at ASTIN Colloquium, Portimao, Portugal.

[7] Mack, T. (1991) A simple parametric model for rating automobile insurance or estimating IBNR claims reserves. *ASTIN Bulletin* **21(1)**, 93-109.

[8] Mack, T. (1993) Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin* **23(2)**, 213-225.

[9] Mack, T. (2008) The prediction error of Bornhuetter/Ferguson. *ASTIN Bulletin* **38(1)**, 87-103.

[10] Neuhaus, W. (1992) Another pragmatic loss reserving method or Bornhuetter/Ferguson revisited. *Scand. Actuarial J.* **2**, 151-162.

[11] Radtke, M. and Schmidt, K.D. (2004) *Handbuch zur Schadenreservierung*. Karlsruhe: Verlag Versicherungswirtschaft, S. 47ff.

[12] Schmidt, K.D. and Wünsche, A. (1998) Chain ladder, marginal sum and maximum likelihood estimation. *Blätter DGVM* **23(3)**, 267-277.

[13] Wüthrich, M.V. and Merz, M. (2008) *Stochastic Claims Reserving Methods in Insurance*. Wiley Finance.

Annina Saluz
ETH Zurich,
RiskLab
Department of Mathematics
8092 Zurich
Switzerland

Alois Gisler
ETH Zurich,
RiskLab
Department of Mathematics
8092 Zurich
Switzerland

Mario V. Wüthrich
ETH Zurich,
RiskLab
Department of Mathematics
8092 Zurich
Switzerland