Chain Ladder Method: Bayesian Bootstrap versus Classical Bootstrap

Gareth W. Peters\textsuperscript{1,2} Mario V. Wüthrich\textsuperscript{3} Pavel V. Shevchenko\textsuperscript{2}

\textsuperscript{1} UNSW Mathematics and Statistics Department, Sydney, 2052, Australia; email: peterga@maths.unsw.edu.au
\textsuperscript{2} CSIRO Sydney, Locked Bag 17, North Ryde, NSW, 1670, Australia
\textsuperscript{3} ETH Zurich, Department of Mathematics, CH-8092 Zurich, Switzerland
Chain Ladder Method: Bayesian Bootstrap versus Classical Bootstrap

Gareth W. Peters\textsuperscript{1,2} Mario V. Wüthrich\textsuperscript{3} Pavel V. Shevchenko\textsuperscript{2}

Abstract

The intention of this paper is to analyse the mean square error of prediction (MSEP) under the distribution-free chain ladder (DFCL) claims reserving method. We compare the estimation obtained from the classical bootstrap method with the one obtained from a Bayesian bootstrap. To achieve this in the DFCL model we develop a novel approximate Bayesian computation (ABC) sampling algorithm to obtain the empirical posterior distribution. We need an ABC sampling algorithm because we work in a distribution-free setting. The use of this ABC methodology combined with bootstrap allows us to obtain samples from the intractable posterior distribution without the requirement of any distributional assumptions. This then enables us to calculate the MSEP and other risk measures like Value-at-Risk.

Key words: Claims reserving, distribution-free chain ladder, mean square error of prediction, Bayesian chain ladder, approximate Bayesian computation, Markov chain Monte Carlo, adaption, annealing, bootstrap
1. Motivation

The distribution-free chain ladder model (DFCL) of Mack [13] is probably the most popular model for stochastic claims reserving. In this paper we use a time series formulation of the DFCL model which allows for bootstrapping the claims reserves. An important aspect of this model is that it can provide a justification for the classical deterministic chain ladder (CL) algorithm which is not founded on an underlying stochastic model. Moreover, it allows for the study of prediction uncertainties. Note that there are different stochastic models that lead to the CL reserves (see for example Wüthrich-Merz [22], Section 3.2). In the present paper we use the DFCL formulation to reproduce the CL reserves.

This paper analyses the parameter estimates in the DFCL model, the associated claims reserves and the mean square errors of prediction (MSEP) from both the frequentist perspective and a contrasting Bayesian view. In doing so we analyse CL point estimators for parameters of the DFCL model, the resulting estimated reserves and the associated MSEP from the classical perspective. These include bootstrap estimated prediction errors obtained via a process of conditional back propagation. These classical frequentist estimators are compared to Bayesian point estimators. The Bayesian estimates considered are the maximum a posteriori (MAP) and the minimum mean square error (MMSE) estimators and the associated Bayesian estimated reserves conditional on such point estimators. In addition, since in the Bayesian setting we obtain samples from the posterior on the parameters we use these along with the MSEP obtained by the point estimators to obtain associated posterior predictive intervals to be compared with the classical bootstrap procedures. We then robustify the prediction of reserves and analyse the resultant MSEP when we integrate out the influence of the prediction of unknown variance parameters in the DFCL model, again achievable since in the Bayesian setting we obtain samples from the joint posterior for the CL factors and the variances.

This requires that we develop a Bayesian CL model for the DFCL model which makes no parametric assumptions on the form of the likelihood function, see also Gisler-Wüthrich [11]. This is unlike the works of Yao [23] and Peters-Shevchenko-Wüthrich [19] that assume explicit distributions in order to construct the posterior distributions in the Bayesian context. Instead we demonstrate how to work directly with the intractable likelihood functions and the resulting intractable posterior distribution, using appropriate distance functions. In this regard we demonstrate that we do not need to make any parametric assumptions to perform posterior inference, avoiding the often poor model assumptions made, as for example in the paper of Yao.
To achieve this we develop a novel approximate Bayesian computation (ABC) sampling algorithm to obtain samples from the posterior. We develop the first use of ABC methodology combined with bootstrap which allows us to obtain samples from the intractable posterior distribution without the requirement of any distributional assumptions or even the requirement on generating of observations from the intractable likelihood model. Instead, we simply require that certain model assumptions are not violated in our conditional back propagation bootstrap procedure. In this context we analyse several aspects of the ABC methodology, including sensitivity studies and convergence diagnostics, before we make use of the algorithm to perform a detailed comparison with classical models.

Outline of this paper. The paper begins with a presentation of the claims reserving problem and then presents the model we shall consider. This is followed by the description of the classical chain ladder algorithm and the construction of a Bayesian model that can be used to estimate the parameters of the model. The Bayesian model is constructed in a distribution-free setting. Following this is a discussion on classical versus Bayesian parameter estimators along with a bootstrap based procedure for estimation of parameter uncertainty in the classical setting. The next section presents the methodology of ABC coupled with a novel bootstrap based sampling procedure which will allow us to work directly with the distribution-free Bayesian model. We then illustrate the developed algorithm on a synthetic data set and the real data set, comparing performance to the classical results and those obtained via credibility theory.

2. Claims development triangle and DFCL model

We briefly outline the claims development triangle structure we utilise in formulation of our models. Assume there is a run-off triangle containing claims development data with the structure given in Table 1.

Assume that \( C_{i,j} \) are cumulative claims with indices \( i \in \{0, \ldots, I\} \) and \( j \in \{0, \ldots, J\} \), where \( i \) denotes the accident year and \( j \) denotes the development year (cumulative claims can refer to payments, claims incurred, etc). We make the simplifying assumption that the number of accident years is equal to the number of observed development periods i.e. \( I = J \). At time \( I \), we have observations

\[
D_I = \{C_{i,j}; \ i + j \leq I\},
\]  

(2.1)
and for claims reserving at time $I$ we need to predict the future claims

$$D_I^c = \{C_{i,j}; \ i + j > I, \ i \leq I, j \leq J\}. \quad (2.2)$$

Moreover, we define for $j \in \{0, \ldots, I\}$ the set $B_j = \{C_{i,k}; \ i + k \leq I, 0 \leq k \leq j\}$, e.g. $B_0$ is the first column in Table 1.

### 2.1. Classical chain ladder algorithm

In the classical (deterministic) chain ladder algorithm there is no underlying stochastic model. It is rather a recursive algorithm that was used to estimate the claims reserves and which has proved to give good practical results. It simply involves the following recursive steps to predict unobserved cumulative claims in $D_I^c$. Set $\hat{C}_{i,I-i} = C_{i,I-i}$ and for $j > I-i$

$$\hat{C}_{i,j} = \hat{C}_{i,j-1}\hat{f}_{j-1}^{(CL)} \quad \text{with CL factor estimates} \quad \hat{f}_{j-1}^{(CL)} = \frac{\sum_{i=0}^{I-j} C_{i,j}}{\sum_{i=0}^{I-j} C_{i,j-1}}. \quad (2.3)$$

Since this is a deterministic algorithm it does not allow for quantification of the uncertainty associated with the predicted reserves. To analyse the associated uncertainty there are several stochastic models that reproduce the CL reserves: for example Mack’s distribution-free chain ladder model [13], the over-dispersed Poisson model (see England-Verrall [5]) or the Bayesian chain ladder model (see Gisler-Wüthrich [11]). We use a time series formulation of the Bayesian chain ladder model in order to use bootstrap methods and Bayesian inference.

### 2.2. Bayesian DFCL model

We use an additive time series version of the Bayes chain ladder model (Model Assumptions 3.1, in Gisler-Wüthrich [11]).
Model Assumptions 2.1.

1. We define the CL factors by \( F = (F_0, \ldots, F_{J-1}) \) and the standard deviation parameters by \( \Xi = (\Xi_0, \ldots, \Xi_{J-1}) \). We assume independence between all these parameters, i.e., the prior density of \((F, \Xi)\) is given by

\[
\pi(f, \sigma) = \prod_{j=0}^{J-1} \pi(f_j) \pi(\sigma_j),
\]

(2.4)

where \( \pi(f_j) \) denotes the density of \( F_j \) and \( \pi(\sigma_j) \) denotes the density of \( \Xi_j \).

2. Conditionally, given \( F = f = (f_0, \ldots, f_{J-1}) \) and \( \Xi = \sigma = (\sigma_0, \ldots, \sigma_{J-1}) \), we have:

- Cumulative claims \( C_{i,j} \) in different accident years \( i \) are independent.
- Cumulative claims satisfy the following time series representation

\[
C_{i,j+1} = f_j C_{i,j} + \sigma_j \sqrt{C_{i,j}} \epsilon_{i,j+1},
\]

(2.5)

where conditionally, given \( B_0 \), we have that the residuals \( \epsilon_{i,j} \) are i.i.d. satisfying

\[
E[\epsilon_{i,j}|B_0, F, \Xi] = 0 \quad \text{and} \quad \text{Var}[\epsilon_{i,j}|B_0, F, \Xi] = 1,
\]

(2.6)

and \( P[C_{i,j} > 0| B_0, F, \Xi] = 1 \) for all \( i, j \).

Remark. Note that the assumptions on the residuals are slightly involved in order to guarantee that cumulative claims \( C_{i,j} \) are positive \( P\text{-a.s.} \).

Corollary 2.2. Under Model Assumptions 2.1 we have that conditionally, given \( D_I \), the random variables \( F_0, \ldots, F_{J-1}, \Xi_0, \ldots, \Xi_{J-1} \) are independent.

Proof of Corollary 2.2. The proof is completely analogous to the proof of Theorem 3.2 in Gisler-Wüthrich [11] and follows from prior independence of the parameters and the fact that \( C_{i,j+1} \) only depends on \( F_j \), \( \Xi_j \) and \( C_{i,j} \) (Markov property).

\[ \square \]

In particular, Corollary 2.2 says that we obtain the following posterior distribution for \((F, \Xi)\), given \( D_I \),

\[
\pi (f, \sigma|D_I) = \prod_{j=0}^{J-1} \pi (f_j|D_I) \pi (\sigma_j|D_I).
\]

(2.7)

This has important implications for the ABC sampling algorithm developed below.

In order to perform the Bayesian analysis we make explicit assumptions on the prior distributions of \((F, \Xi)\).
Model Assumptions 2.3.

In addition to Model Assumptions 2.1 we assume that the prior model for all parameters $j \in \{0, \ldots, J - 1\}$ is given by:

- $F_j \sim \Gamma (\alpha_j, \beta_j)$, where $\Gamma (\alpha_j, \beta_j)$ is a gamma distribution with mean $E[F_j] = \alpha_j \beta_j = \hat{f}_j^{(CL)}$ (see (2.3)) and large variance to have diffuse priors.

- The variances $\Xi_j^2 \sim IG (a_j, b_j)$, where $IG (a_j, b_j)$ is an inverse gamma distribution with mean $E[\Xi_j^2] = b_j / (a_j - 1) = \hat{\sigma}_j^{2(CL)}$ (see (2.8), below) and large variance.

Remarks.

1) The likelihood model is intractable, meaning that no density can be written down analytically in the DFCL model. We have only made distributional assumptions on the parameters $(F, \Xi)$ but not on the observable cumulative claims $C_{i,j}$. Therefore, a full Bayesian analysis using analytic posterior distributions cannot be performed. One way out of this dilemma would be to formulate a full Bayesian model by making distributional assumptions (this is, e.g., done in Yao [23]) but then the model is no longer distribution-free. Another approach would be to use credibility methods (see Gisler-Wüthrich [11]) but this only gives statements for second moments. In the present set up we develop ABC methods that allow for a full distributional answer for the posterior without making explicit distributional assumptions for the cumulative claims $C_{i,j}$.

2) We select the priors to ensure that we maintain several relevant aspects of the DFCL model. Firstly, it is important to utilise priors that enforce the strict positivity of the parameters $f_j, \sigma_j > 0$. We note here that the parametric Bayesian model developed in Yao [23] failed in this aspect and therefore we develop an alternative prior structure that satisfies these required properties of the DFCL model. Secondly, our priors are chosen such that they do not contain information, i.e. we assume diffuse priors with large variances.

2.3. Classical and Bayesian parameter estimators

In the classical CL method, the CL factors are estimated by $\hat{f}_j^{(CL)}$ given in (2.3). The variance parameters are estimated by, see e.g. (3.4) Wüthrich-Merz [22],

$$
\hat{\sigma}_j^{2(CL)} = \frac{1}{I - j - 1} \sum_{i=0}^{I-j-1} C_{i,j} \left( \frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j^{(CL)} \right)^2.
$$

(2.8)

Note that this estimator is only well-defined for $j < I - 1$. There is a vast literature and discussion on the estimation of tail parameters. We do not enter this discussion here but we
simply choose the estimator given in Mack [13] for the last variance parameter which is defined by
\[ \hat{\sigma}_{J-1}^{2(CL)} = \min \left\{ \frac{\hat{\sigma}_{J-2}^{2(CL)}}{\hat{\sigma}_{J-3}^{2(CL)}}, \hat{\sigma}_{J-3}^{2(CL)} \right\}. \] (2.9)

In a Bayesian inference context one calculates the posterior distribution of the parameters, given \( D_I \). As in (2.7) we denote this posterior by \( \pi(f, \sigma | D_I) \). Then, there are two commonly used point estimators in Bayesian analysis that correspond to the posterior mode (MAP) and the posterior mean (MMSE), respectively. Given the posterior independence (see Corollary 2.2) they are given by:
\[ \hat{f}_j^{(MAP)} = \arg \max_{f_j} \pi(f_j | D_I), \] (2.10)
\[ \hat{\sigma}_j^{(MAP)} = \arg \max_{\sigma_j} \pi(\sigma_j | D_I), \] (2.11)
and
\[ \hat{f}_j^{(MMSE)} = \int f_j \pi(f_j | D_I) df_j = E[f_j | D_I], \] (2.12)
\[ \hat{\sigma}_j^{(MMSE)} = \int \sigma_j \pi(\sigma_j | D_I) d\sigma_j = E[\sigma_j | D_I]. \] (2.13)

Note that for diffuse prior we find (see Corollary 5.1 in Gisler-Wüthrich [11])
\[ \hat{f}_j^{(MMSE)} \approx \hat{f}_j^{(CL)}. \] (2.14)

Hence, using Corollary 2.2, we obtain the approximation
\[ E[C_{i,j} | D_I] = E[E[C_{i,j} | D_I, F, \Xi] | D_I] = C_{i,I-i} E\left[ \prod_{j=I-i}^{J-1} F_j | D_I \right] \]
\[ \approx C_{i,I-i} \prod_{j=I-i}^{J-1} \hat{f}_j^{(CL)} = \hat{C}_{i,j}, \] (2.15)
where on the last line we have an asymptotic equality if the diffusivity of the priors \( \pi(f_j) \) tends to infinity. This is exactly the argument why the Bayesian CL model can be used to justify the CL predictors, see Gisler-Wüthrich [11].

3. Bootstrap and mean square error of prediction

Assume that we have calculated the Bayesian predictor or the CL predictor given in (2.15). Then we would like to determine the prediction uncertainty, i.e., we would like to study the
deviation of \( C_{i,J} \) around its predictor. If one is only interested in second moments, the so-called conditional mean square error of prediction (MSEP), one can often estimate the error terms analytically. However, other uncertainty measures like Value-at-Risk (VaR) can only be determined numerically.

A popular numerical method is the bootstrap method. The bootstrap technique was developed by Efron [3] and extended by Efron-Tibshirani [4] and Davison-Hinkley [1]. This procedure allows one to obtain information regarding an aggregated distribution given a single realisation of the data. To apply the bootstrap procedure one introduces a minimal amount of model structure such that resampling observations can be achieved using observed samples of the data. In this section we present a bootstrap algorithm in the classical frequentists approach, i.e., we assume that the CL factors \( F = f \) and the standard deviation parameters \( \Xi = \sigma \) given in Model Assumptions 2.1 are unknown constants. The bootstrap then generates synthetic data denoted by \( D^*_I \) that allow for the study of the fluctuations of \( \hat{f}^{(CL)}(\text{CL}) \) and \( \hat{\sigma}^2(\text{CL}) \) (for details see Wüthrich-Merz [22], Section 7.4). In the presented text we restrict ourselves to the conditional resampling approach presented in Section 7.4.2 of Wüthrich-Merz [22].

---

**Non-parametric classical bootstrap (conditional version).**

1. Calculate estimated residuals \( \tilde{\varepsilon}_{i,j} \) for \( i + j \leq I, j > 0 \), conditional on the estimators \( \hat{f}_{0:j-1}^{(CL)} \) and \( \hat{\sigma}^2_{0:j-1}^{(CL)} \) and the observed data \( D_I \):

\[
\tilde{\varepsilon}_{i,j} = \tilde{\varepsilon}_{i,j}(\hat{f}_{j-1}^{(CL)}, \hat{\sigma}_{j-1}^{(CL)}) = \frac{C_{i,j} - \hat{f}_{j-1}^{(CL)} C_{i,\hat{j}-1}}{\hat{\sigma}_{j-1}^{(CL)} \sqrt{C_{i,\hat{j}-1}}}.
\]

2. These residuals \( (\tilde{\varepsilon}_{i,j})_{i+j\leq I} \) give the empirical bootstrap distribution \( \hat{F}_{D_I} \).

3. Sample i.i.d. residuals \( \tilde{\varepsilon}_{i,j}^* \sim \hat{F}_{D_I} \) for \( i + j \leq I, j > 0 \).

4. Generate bootstrap observations (conditional resampling)

\[
C^*_{i,j} = \hat{f}_{j-1}^{(CL)} C_{i,j-1} + \hat{\sigma}_{j-1}^{(CL)} \sqrt{C_{i,j-1} \tilde{\varepsilon}_{i,j}^*},
\]

which defines \( D_I^* = D_I^*(\hat{f}^{(CL)}, \hat{\sigma}^{(CL)}) \). Note that for the unconditional version of bootstrap we should generate \( C^*_{i,j} = \hat{f}_{j-1}^{(CL)} C_{i,j-1} + \hat{\sigma}_{j-1}^{(CL)} \sqrt{C_{i,j-1} \tilde{\varepsilon}_{i,j}^*} \). For a discussion on this approach see Section 7.4.1 of [22].

5. Calculate bootstrapped CL parameters \( \hat{f}^*_j \) and \( \hat{\sigma}^2_\hat{j} \) by

\[
\hat{f}^*_j = \frac{\sum_{i=0}^{I-j-1} C^*_{i,j+1}}{\sum_{i=0}^{I-j-1} C_{i,j}},
\]

\[
\hat{\sigma}^2_\hat{j} = \frac{1}{I-j-1} \sum_{i=0}^{I-j-1} C_{i,j} \left( \frac{C^*_{i,j+1}}{C_{i,j}} - \hat{f}^*_j \right)^2.
\]
6. Repeat steps 3-5 and obtain empirical distributions from the bootstrap samples \( \hat{C}_{i,j}^* \), \( \hat{f}_j^* \) and \( \hat{\sigma}_j^2 \). These are then used to quantify the parameter estimation uncertainty.

This non-parametric classical bootstrap method can be seen as a frequentist approach. This means that we do not express our parameter uncertainty by the choice of an appropriate prior distribution. We rather use a point estimator for the unknown parameters and then study the possible fluctuations of this point estimator.

The main difficulty now is that the non-parametric bootstrap method, as described above, underestimates the “true” uncertainty. This comes from the fact that the estimated residuals \( \tilde{\varepsilon}_{i,j} \), in general, have variance smaller than 1 (see formula (7.23) in Wüthrich-Merz [22]). This means that our estimated residuals are not appropriately scaled. Therefore, frequentists use several different scalings to correct this fact (see formula (7.24) in Wüthrich-Merz [22] or England-Verrall [5]). Here, we use a different approach by introducing the Bayesian bootstrap method, see Section 4 below.

**Frequentist bootstrap estimates.**

Let us for the time-being concentrate on the conditional MSEP given by

\[
\text{msep}_{C_{i,j}|D_I} \left( \hat{C}_{i,j} \right) = E \left[ \left( C_{i,j} - \hat{C}_{i,j} \right)^2 \bigg| D_I \right] = \text{Var} \left( C_{i,j} | D_I \right) + \left( E \left[ C_{i,j} | D_I \right] - \hat{C}_{i,j} \right)^2.
\]

(3.1)

The first term is known as the conditional process variance and the second term as the parameter estimation uncertainty. In the frequentists approach, i.e. for given deterministic \( F = f \) and \( \Xi = \sigma \), this can be calculated, see Wüthrich-Merz [22], Section 3.2. Namely, the terms are given by

\[
\text{Var} \left( C_{i,j} | D_I \right) = \left( E \left[ C_{i,j} | C_{i,I-i} \right] \right)^2 \sum_{j=I-i}^{J-1} \frac{\hat{\sigma}_j^2 \hat{f}_j^2}{E \left[ C_{i,j} | C_{i,I-I} \right]^2} \overset{def.}{=} C_{i,I-i} \Gamma_{I-i},
\]

(3.2)

and

\[
\left( E \left[ C_{i,j} | D_I \right] - \hat{C}_{i,j} \right)^2 = C_{i,I-I-i}^2 \left( \prod_{j=I-i}^{J-1} f_j - \prod_{j=I-I-i}^{J-1} \tilde{f}_j^{(CL)} \right)^2 \overset{def.}{=} C_{i,I-I-i}^2 \Delta_{I-i}.
\]

(3.3)

The process variance (3.2) is estimated by replacing the parameters by its estimators,

\[
\tilde{\text{Var}} \left( C_{i,j} | D_I \right) = \left( \hat{C}_{i,j} \right)^2 \sum_{j=I-I-i}^{J-1} \frac{\hat{\sigma}_j^{2(CL)} / (\tilde{f}_j^{(CL)})^2}{\hat{C}_{i,j}} \overset{def.}{=} C_{i,I-I-i} \hat{\Gamma}_{I-I-i}^\text{freq}.
\]

(3.4)

10
The parameter estimation error is more involved and there we need the bootstrap algorithm. Assume that the bootstrap method gives $T$ bootstrap samples $\hat{f}_j^{(1)}, \ldots, \hat{f}_j^{(T)}$. Then the parameter estimation error (3.3) is estimated by the sample variance of the product of the bootstrap observation chain ladder parameter estimates $\hat{f}_j^{(1)}, \ldots, \hat{f}_j^{(T)}$, which gives the estimator $C_{i,i-1}^2 \hat{\Delta}_{i-i}^{\text{freq}}$.

**Bayesian estimates.**

In the Bayesian setup, i.e. choosing prior distributions for the unknown parameters $F$ and $\Xi$, we obtain a natural decomposition of the conditional MSEP.

$$\text{mse}_{C_{i,j}|D_I} (E [C_{i,j} | D_I]) = \text{Var} (C_{i,j} | D_I)$$

$$= E \left[ \text{Var} (C_{i,j} | D_I, F, \Xi) | D_I \right] + \text{Var} (E [C_{i,j} | D_I, F, \Xi] | D_I).$$

The average process variance is given by (see Wüthrich-Merz [22], Lemma 3.6)

$$E \left[ \text{Var} (C_{i,j} | D_I, F, \Xi) | D_I \right] = C_{i,I-i} \sum_{j=1-i}^{J-1} \prod_{m=I-i}^{j-1} F_m \Xi^2 \prod_{n=j+1}^{J-1} F_n^2 | D_I$$

$$= C_{i,I-i} \prod_{m=I-i}^{J-1} \prod_{n=j+1}^{J-1} E \left[ F_m | D_I \right] E \left[ \Xi^2 | D_I \right] \prod_{n=j+1}^{J-1} E \left[ F_n^2 | D_I \right] \overset{\text{def.}}{=} C_{i,I-i} \tilde{\Delta}_{i-i}^{\text{Bayes}},$$

where we have used Corollary 2.2. The parameter estimation error is given by

$$\text{Var} (E [C_{i,j} | D_I, F, \Xi] | D_I) = C_{i,I-i}^2 \text{Var} \left( \prod_{j=I-i}^{J-1} F_j | D_I \right) \overset{\text{def.}}{=} C_{i,I-i}^2 \tilde{\Delta}_{i-i}^{\text{Bayes}},$$

where we have used (2.15). Using posterior independence, see Corollary 2.2, we obtain for the last term

$$C_{i,I-i}^2 \tilde{\Delta}_{i-i}^{\text{Bayes}} = C_{i,I-i}^2 \left[ \prod_{j=I-i}^{J-1} E \left[ F_j^2 | D_I \right] - \prod_{j=I-i}^{J-1} E \left[ F_j | D_I \right]^2 \right].$$

In order to calculate these two terms given in (3.6) and (3.8) we need to calculate the posterior distribution of $(F, \Xi)$, given $D_I$. Since we do not have a full distributional model, we cannot write down the likelihood function, which would allow for analytical solutions or Markov chain Monte Carlo (MCMC) simulations. Therefore we introduce the ABC framework which allows for distribution-free simulations using appropriate bootstrap samples and a distance metric. This we are going to discuss in the next section.

**Credibility Estimates.**

As mentioned previously, we can also consider the credibility estimates given in Gisler-Wüthrich
As long as we are only interested in the second moments, i.e. conditional MSEP, we can also use credibility estimators, which are minimum variance estimators that are linear in the observations. For diffuse priors we obtain the approximation given in Corollary 7.2 of Gisler-Wüthrich

\[
\hat{\text{mse}}_{P|D}(E[C_{i,j}|D_I]) = C_{i,I-i} \hat{\Gamma}^{\text{cred}}_{I-i} + C_{i,i}^2 \hat{\Delta}^{\text{cred}}_{I-i},
\]

where

\[
\hat{\Gamma}^{\text{cred}}_{I-i} = \sum_{j=I}^{J-1} \left\{ \prod_{m=I-i}^{j-1} \hat{f}_{m}^{(CL)} \hat{\sigma}_{j}^{2(\text{CL})} \prod_{n=j+1}^{J-1} \left( (\hat{f}_{n}^{(CL)})^2 + \frac{\hat{\sigma}_{n}^{2(\text{CL})}}{\sum_{i=0}^{I-n-1} C_{i,n}} \right) \right\}, \quad (3.10)
\]

\[
\hat{\Delta}^{\text{cred}}_{I-i} = \prod_{j=I-i}^{J-1} \left( (\hat{f}_{j}^{(CL)})^2 + \frac{\hat{\sigma}_{j}^{2(\text{CL})}}{\sum_{i=0}^{I-j-1} C_{i,j}} \right) - \prod_{j=I-i}^{J-1} (\hat{f}_{j}^{(CL)})^2. \quad (3.11)
\]

In the results section we compare the frequentist bootstrap approach, the credibility approach and the ABC bootstrap approach that is described below (see Table 7 below).

Note that the ABC bootstrap approach can also be applied to other risk measures, not just MSEP. This is unlike the credibility approach which only applies to MSEP estimates. For example, if we fix a security level 95% we can ask for the VaR on that level, which is defined by

\[
\text{VaR}_{0.95}\left( C_{i,j} - E[C_{i,j}|D_I] \bigg| D_I \right) = \min \left\{ x; \ P\left[ C_{i,j} - E[C_{i,j}|D_I] > x \bigg| D_I \right] \leq 0.05 \right\}. \quad (3.12)
\]

4. ABC for intractable likelihoods

Obtaining samples \( \{ f^{(t)}, \sigma^{2(t)} \}_{t=1:T} \) which are realisations of a random vector distributed with a posterior distribution \( \pi(f, \sigma|D) \) in the DFCL model is difficult since the likelihood is intractable. Hence, standard approaches such as Markov chain Monte Carlo (MCMC) algorithms (see, e.g., Gilks et al. [10]) cannot be directly used. Other authors have avoided this difficulty by making distributional assumptions for the form of the likelihood. This then violates a distribution-free CL model assumption but allows for relatively standard sampling procedures to be applied. Note that a specific Gaussian assumption does not allow for skewness. Here, we do not make any such assumptions and instead we work in a truly distribution-free model using approximate Bayesian computation (ABC) to facilitate sampling from an intractable posterior distribution. To sample from the posterior in our DFCL model we develop a novel formulation of the ABC methodology based on the bootstrap and conditional back transformation procedure, similar to that discussed in Section 3.

ABC methods aim to sample from posterior distributions in the presence of computationally intractable likelihood functions. For an application in risk modelling of ABC techniques see
Peters-Sisson [15]. In this article we present a novel ABC-MCMC algorithm. Alternatively, sequential Monte Carlo (SMC) based algorithms which can improve simulation efficiency can be found in Del Moral et al. [2], Sisson et al. [21], Peters et al. [16],[17] and Marjoram et al. [14]. These algorithms are beyond the scope of the present paper.

We provide a basic argument for how the ABC methodology works in Appendix A. For a given observation \( y \) we want to sample from \( \pi(\theta|y) \) with an intractable likelihood function. We assume that \( S(y) \) is a sufficient statistic for the model from which we assume data \( y \) is a realisation.

We define a hard decision function \( g(x, y) = \mathbb{I}\{\rho(S(x), S(y)) < \epsilon\}(x) \) for a given tolerance level \( \epsilon > 0 \) and a distance metric \( \rho(\cdot, \cdot) \), where \( \mathbb{I}\{\cdot\} \) is the indicator function which equals 1 if the event is true and 0 otherwise. As demonstrated in Appendix A we use the approximation, see (A.3)-(A.4),

\[
\begin{align*}
\pi(\theta|y) &\approx \frac{\int g(x, y) \pi(\theta) \pi(x|\theta) \, dx}{\int g(x, y) \pi(\theta) \pi(x|\theta) \, dx d\theta} = \frac{\pi(\theta) E[g(X, y)|\theta]}{E[g(X, y)]}, \\
\end{align*}
\]

where \( X \sim \pi(x|\theta) \) and \( \theta \sim \pi(\theta) \). In the next step the numerator of (4.1) is approximated using the empirical distribution, i.e.

\[
\pi(\theta) E[g(X, y)|\theta] \approx \pi(\theta) \frac{1}{L} \sum_{l=1}^{L} g\left(x^{(l)}(\theta), y\right),
\]

where \( x^{(l)}(\theta) \overset{i.i.d.}{\sim} \pi(x|\theta) \). Finally, we need to consider the denominator \( E[g(X, y)] \). In general this has a non-trivial form that cannot be calculated analytically. However, since we use an MCMC based method the denominators cancel in the accept-reject stage of the algorithm. Therefore the intractability of the denominator does not impede sampling from the posterior. Therefore, we use

\[
\begin{align*}
\pi(\theta|y) &\approx \frac{\int g(x, y) \pi(\theta) \pi(x|\theta) \, dx}{\int g(x, y) \pi(\theta) \pi(x|\theta) \, dx d\theta} \propto \pi(\theta) E[g(X, y)|\theta] \approx \pi(\theta) \frac{1}{L} \sum_{l=1}^{L} g\left(x^{(l)}(\theta), y\right) \approx \pi(\theta) \frac{1}{L} \sum_{l=1}^{L} g\left(x^{(l)}(\theta), y\right),
\end{align*}
\]

in order to obtain samples from \( \pi(\theta|y) \). Almost universally, \( L = 1 \) is adopted to reduce computation but on the other hand this will slow down the rate of convergence to the stationary distribution.

Note that sometimes one also uses softer decision functions for \( g(x, y) \). The role of the distance measure \( \rho \) is evaluated by Fan et al. [6], we further extend this analysis to the class of models considered in this paper. We analyse several choices for the distance measure \( \rho \) such as Mahalanobis distance, scaled Euclidean distance and the Manhattan ”City Block” distance. Fan et al. [6] demonstrate that it is not efficient to utilise the standard Euclidean distance, especially when summary statistics considered are on different scales.
Additionally, using an ABC-MCMC algorithm, it is important to assess convergence diagnostics. Particularly when using ABC-MCMC where serial correlation in the Markov chain samples can be significant if the sampler is not designed carefully. We assess autocorrelation of the simulated Markov chain, the Geweke [9] time series statistic and the Gelman-Rubin [7] R-statistic convergence diagnostic in an ABC setting.

Concluding: We apply three different techniques in order to treat the intractable likelihood: (1) we use ABC to get a handle on the likelihood, (2) therefore we need synthetic samples from the DFCL model given realisations of the parameters, these come from the bootstrap algorithm and (3) by using a well understood MCMC based sampling algorithm we obtain cancellation of the non-analytic normalizing constants. Next we discuss the ABC-MCMC algorithm in detail.

4.1. ABC algorithmic choices for the time series DFCL model

We start with the choices of the ABC components.

- **Data set:** The observations are given by the cumulative claims denoted by $D_I$.

- **Generation of a synthetic data set:** Note that in this setting not only is the likelihood intractable but additionally the generation of a synthetic data set $D_I^*$ given the current parameter values $F, \Xi$ is not straightforward. The synthetic data set $D_I^*$ is generated using the bootstrap procedure described in Section 3. Note that both the bootstrap residual $\tilde{\epsilon}_{i,j}$ and the bootstrap samples $D_I^*$ are functions of the parameter choices, see page 9. Therefore we generate for given $F = f$ and $\Xi = \sigma$ the bootstrap residuals $\tilde{\epsilon}_{i,j} = \tilde{\epsilon}_{i,j}(f_{j-1}, \sigma_{j-1})$ and the bootstrap samples $D_I^* = D_I^*(f, \sigma)$ according to the non-parametric bootstrap (see page 9) where we replace the CL parameter estimates $(\hat{f}^{(CL)}, \hat{\sigma}^{(CL)})$ by the parameters $\theta = (F, \Xi)$.

- **Summary statistics:** In order to define the decision function $g$ we introduce summary statistics, see Appendix A for details of the role these summary statistics play in the ABC approximation. For the observed data $D_I$ we define the vector

$$S (D_I; 0, 1) = (S_1, \ldots, S_{n+2})$$

$$= (C_{0,1}, \ldots, C_{0,J}, C_{1,1}, \ldots, C_{0,J-1}, \ldots, C_{I-2,1}, C_{I-2,2}, C_{I-1,1}; 0, 1),$$

where $n$ denotes the number of residuals $\tilde{\epsilon}_{i,j}$. For given $\theta = (F, \Xi)$ we generate the bootstrap sample $D_I^* = D_I^*(F, \Xi)$ as described above. The corresponding residuals $\tilde{\epsilon}_{i,j} = \tilde{\epsilon}_{i,j}$.
$\bar{\varepsilon}_{i,j}(F_{j-1},\Xi_{j-1})$ should also be close to the standardized observations. Therefore we define its empirical mean and standard deviation by

$$
\mu^* = \mu^*(F,\Xi) = \frac{1}{n} \sum_{i,j} \bar{\varepsilon}_{i,j}(F_{j-1},\Xi_{j-1}), \quad (4.4)
$$

$$
s^* = s^*(F,\Xi) = \left[ \frac{1}{n-1} \sum_{i,j} (\bar{\varepsilon}_{i,j}(F_{j-1},\Xi_{j-1}) - \mu^*(F,\Xi))^2 \right]^{1/2}. \quad (4.5)
$$

Hence, the summary statistics for the synthetic data is given by

$$
S(D^*; \mu^*, s^*) = (C^*_{0,1}, \ldots, C^*_{0,J}, C^*_{1,1}, \ldots, C^*_{0,J-1}, \ldots, C^*_{I-2,1}, C^*_{I-2,2}, C^*_{I-1,1}, \mu^*, s^*).$$

**Distance metrics:**

- **Mahalanobis distance and scaled Euclidean distance**
  Here we draw on the analysis of Sisson et al. [6] who propose the use of the Mahalanobis distance metric given by,

$$
\rho(S(D_I; 0, 1), S(D^*_I; \mu^*, s^*)) = [S(D_I; 0, 1) - S(D^*_I; \mu^*, s^*)]^{\top} \Sigma_{D_I}^{-1} [S(D_I; 0, 1) - S(D^*_I; \mu^*, s^*)],
$$

where the covariance matrix $\Sigma_{D_I}$ is an appropriate scaling described in Appendix B. The scaled Euclidean distance is obtained when we only consider the diagonal elements of the covariance matrix $\Sigma_{D_I}$.

Note, the covariance matrix $\Sigma_{D_I}$ provides a weighting on each element of the vector of summary statistics to ensure they are scaled appropriately according to their influence on the ABC approximation. There are many other such weighting schemes one could conceive.

- **Manhattan "City Block" distance**
  We consider the $L^1$-distance given by

$$
\rho(S(D_I; 0, 1), S(D^*_I; \mu^*, s^*)) = \sum_{i=1}^{n+2} |S_i(D_I; 0, 1) - S_i(D^*_I; \mu^*, s^*)|. 
$$

**Decision function:** We work with a hard decision function given by

$$
g(D_I, D^*_I) = \mathbb{I}\{\rho(S(D_I; 0, 1), S(D^*_I; \mu^*, s^*)) < \epsilon\}. 
$$
• **Tolerance schedule:** We use the sequence

\[ \epsilon_t = \max \{ 20,000 - 10t, \epsilon_{\min} \} . \]

Note, the use of an ABC-MCMC algorithm can result in "sticking" of the chain for extended periods. Therefore, one should carefully monitor convergence diagnostics of the resulting Markov chain for a given tolerance schedule. There is a trade-off between the length of the Markov chain required for samples approximately from the stationary distribution and the bias introduced by non zero tolerance. In this paper we set \( \epsilon_{\min} \) via preliminary analysis of the Markov chain sampler mixing rates for a transition kernel with coefficient of variation set to one.

We note that in general practitioners will have a required precision in posterior estimates, this precision can be directly used to determine, for a given computational budget, a suitable tolerance \( \epsilon_{\min} \).

• **Convergence diagnostics:** We stress that when using an ABC-MCMC algorithm, it is crucial to carefully monitor the convergence diagnostics of the Markov chain. This is more important in the ABC context than in the general MCMC context due to the possibility of extended rejections where the Markov chain can stick in a given state for long periods. This can be combatted in several ways which will be discussed once the algorithm is presented.

The convergence diagnostics we consider are evaluated only on samples post annealing of the tolerance threshold and after an initial burnin period once tolerance of \( \epsilon_{\min} \) is reached. If the total chain has length \( T \), the initial burnin stage will correspond to the first \( T_b \) samples and we define \( \tilde{T} = T - T_b \). We denote by \( \{\theta_i^{(t)}\}_{t=1:\tilde{T}} \) the Markov chain of the \( i \)-th parameter after burnin. The diagnostics we consider are given by,

- **Autocorrelation.** This convergence diagnostic will monitor serial correlation in the Markov chain. For given Markov chain samples for the \( i \)-th parameter \( \{\theta_i^{(t)}\}_{t=1:\tilde{T}} \) we define the biased autocorrelation estimate at lag \( \tau \) by

\[
\hat{ACF}(\theta_i, \tau) = \frac{1}{(T - \tau)\hat{\sigma}(\theta_i)} \sum_{t=1}^{T-\tau} [\theta_i^{(t)} - \hat{\mu}(\theta_i)][\theta_i^{(t+\tau)} - \hat{\mu}(\theta_i)], \quad (4.6)
\]

where \( \hat{\mu}(\theta_i) \) and \( \hat{\sigma}(\theta_i) \) are the estimated mean and standard deviation of \( \theta_i \).

- **Geweke [9] time series diagnostic.** For parameter \( \theta_i \) we calculate:
1. Split the Markov chain samples into two sequences, \( \{ \theta_i^{(l)} \}_{t=1:T_l} \) and \( \{ \theta_i^{(l)} \}_{t=T^*+1:} \), such that \( T^* = \hat{T} - T_2 + 1 \), and with ratios \( T_1/\hat{T} \) and \( T_2/\hat{T} \) fixed such that \( (T_1 + T_2)/\hat{T} < 1 \) for all \( \hat{T} \).

2. Evaluate \( \hat{\mu} \left( \theta_i^{T_1} \right) \) and \( \hat{\mu} \left( \theta_i^{T_2} \right) \) corresponding to the sample means on each subsequence.

3. Evaluate consistent spectral density estimates for each subsequence, at frequency 0, denoted \( \hat{SD}(0; T_1, \theta_i) \) and \( \hat{SD}(0; T_2, \theta_i) \). The spectral density estimator considered in this paper is the classical non-parametric periodogram or power spectral density estimator. We use Welch’s method with a Hanning window, for details see Appendix C.

4. Evaluate convergence diagnostic given by
   \[
   Z_{\hat{T}} = \frac{\hat{\mu}(\theta_i^{T_1}) - \hat{\mu}(\theta_i^{T_2})}{T_1^{-1} \hat{SD}(0; T_1, \theta_i) + T_2^{-1} \hat{SD}(0; T_2, \theta_i)}.
   \]
   For \( \hat{T} \to \infty \) one has according to the central limit theorem that \( Z_{\hat{T}} \to \mathcal{N}(0, 1) \) if the sequence \( \{ \theta_i^{(l)} \}_{t=1:} \) is stationary.

\textit{Gelman-Rubin [7] R-statistic diagnostic.} This approach to convergence analysis requires that one runs multiple parallel independent Markov chains each starting at randomly selected initial starting points, we run 5 chains. For comparison purposes we split the total computational budget of \( \hat{T} \) into \( T_1 = T_2 = \ldots = T_5 = \frac{\hat{T}}{5} \). We compute the convergence diagnostic for parameter \( \theta_i \) with the following steps:

1. Generate 5 independent Markov chain sequences, producing the chains for parameter \( \theta_i \), denoted \( \{ \theta_i^{(l)} \}_{t=1:T_k} \) for \( k \in \{1, \ldots, 5\} \).

2. Calculate the sample means \( \hat{\mu} \left( \theta_i^{T_k} \right) \) for each sequence, and the overall mean \( \hat{\mu} \left( \theta_i^{\hat{T}} \right) \).

3. Calculate the variance of the sequence means,
   \[
   \frac{1}{5} \sum_{k=1}^{5} \left( \hat{\mu} \left( \theta_i^{T_k} \right) - \hat{\mu} \left( \theta_i^{\hat{T}} \right) \right)^2 \overset{df}{=} B_i/T_k.
   \]

4. Calculate the within-sequence variances \( \hat{s}^2 \left( \theta_i^{T_k} \right) \) for each sequence.

5. Calculate the average within-sequence variance, \( \frac{1}{5} \sum_{k=1}^{5} \hat{s}^2 \left( \theta_i^{T_k} \right) \overset{df}{=} W_i \).

6. Estimate the target posterior variance for parameter \( \theta_i \) by the weighted linear combination, \( \tilde{\sigma}^2 \left( \theta_i^{\hat{T}} \right) = \frac{T_k-1}{T_k} W_i + \frac{1}{T_k} B_i \). This estimate is unbiased for samples which are from the stationary distribution. In the case in which not all sub chains have reached stationarity, this overestimates the posterior variance for a finite \( \hat{T} \) but it asymptotically, \( \hat{T} \to \infty \), converges to the posterior variance.
7. Improve on the Gaussian estimate of the target posterior given by
\[ N(\hat{\mu}(\theta_{\tilde{T}}), \hat{\sigma}^2(\theta_{\tilde{T}})) \] by accounting for sampling variability in the estimates of the posterior mean and variance. This can be achieved by making a Student-t approximation with location \( \hat{\mu}(\theta_{\tilde{T}}) \), scale \( \sqrt{\hat{V}_i} \) and degrees of freedom \( df_i \), each given respectively by;
\[
\hat{V}_i = \hat{\sigma}^2(\theta_{\tilde{T}}) + B_i \quad \text{and} \quad df_i = \frac{2\hat{V}_i^2}{\text{Var}(\hat{V}_i)},
\]
where the variance is estimated as,
\[
\text{Var}(\hat{V}_i) = \frac{1}{5} \left( \frac{T_1}{25T_1} \right)^2 \text{Var}(z^2(\theta^T_k)) + \left( \frac{6}{\sqrt{2\tilde{T}}} \right)^2 B_i^2 \quad + \frac{12(T_1 - 1)}{25T_1} \hat{\mu}(\theta_{\tilde{T}}) \text{Cov}(z^2(\theta^T_k), \hat{\mu}(\theta_{\tilde{T}})) \quad - \frac{24(T_1 - 1)}{25T_1} \hat{\mu}(\theta_{\tilde{T}}) \text{Cov}(z^2(\theta^T_k), \hat{\mu}(\theta_{\tilde{T}})).
\]
(4.7)
Note, the covariance terms are estimated empirically using the within sequence estimates of the mean and variance obtained for each sequence.

8. Calculate the convergence diagnostic, \( \sqrt{\hat{R} = \sqrt{\frac{V_{df_i}}{W_i(df_i - 2)}}} \), where as \( \tilde{T} \to \infty \) one can prove that \( \hat{R} \to 1 \). This convergence diagnostic monitors the scale factor by which the current distribution for \( \theta_i \) may be reduced if simulations are continued for \( \tilde{T} \to \infty \).

We begin by presenting the ABC-MCMC based algorithm we propose for sampling from the posterior distribution presented in Section 2.1.

4.2. ABC-MCMC to sample from \( \pi(f, \sigma | D_I) \)

We develop an ABC-MCMC algorithm which has an adaptive proposal mechanism and annealing of the tolerance during burnin of the Markov chain. Having reached the final tolerance post annealing, denoted \( \epsilon^\text{min} \), we utilize the remaining burnin samples to tune the proposal distribution to ensure an acceptance probability between the range of 0.3 and 0.5 is achieved. The optimal acceptance probability when posterior parameters are i.i.d. Gaussian was proven to be at 0.234, see Gelman et al. [8]. Though our problem does not match the required conditions for this proof, it provides a practical guide. To achieve this, we tune the coefficient of variation of the proposal, in our case the shape parameter of the gamma proposal distribution. We impose an additional constraint that the minimum shape parameter value is set at \( \gamma^\text{min}_j \) for \( j \in \{1, \ldots, 2J\} \).

ABC-MCMC algorithm using bootstrap samples.
1. For $t = 0$ initialize the parameter vector randomly, this gives $\theta_{1:2J}^{(0)} = (\hat{f}_{0:J-1}^{(0)}, \sigma_{0:J-1}^{(0)})$. Initialize the proposal shape parameters, $\gamma_j \geq \gamma_j^{\text{min}}$ for all $j \in \{1, \ldots, 2J\}$.

2. For $t = 1, \ldots, T$

   (a) Set $\left(\theta_{1:2J}^{(t)}\right) = \left(\theta_{1:2J}^{(t-1)}\right)$

   (b) For $j = 1, \ldots, 2J$

   i. Sample proposal $\theta_j^*$ from a $\Gamma(\gamma_j, \theta_j^{(t)}/\gamma_j)$-distribution. We denote the gamma proposal density by $K(\theta_j^*; \gamma_j, \theta_j^{(t)}/\gamma_j)$. This gives proposed parameter vector $\theta^* = \left(\theta_{1:j-1}^{(t)}, \theta_j^*, \theta_{j+1:2J}^{(t)}\right)$.

   ii. Conditional on $\theta^* = \left(\theta_{1:j-1}^{(t)}, \theta_j^*, \theta_{j+1:2J}^{(t)}\right)$, generate synthetic bootstrap data set $D_I^* = D_I^*(\theta^*)$ using the bootstrap procedure detailed in Section 3 where we replace the CL parameter estimates $(\hat{f}_{CL}, \hat{\sigma}_{CL})$ by the parameters $\theta^*$.

   iii. Evaluate summary statistics $S(D_I; 0, 1)$ and $S(D_I^*; \mu^*, s^*)$ and corresponding decision function $g(D_I, D_I^*)$ as described in Section 4.1.

   iv. Accept proposal with ABC acceptance probability

   $$A\left(\theta_{1:2J}^{(t)}, \theta^*\right) = \min\left\{1, \frac{\pi(\theta_j^*) K(\theta_j^*; \gamma_j, \theta_j^{(t)}/\gamma_j)}{\pi(\theta_j^{(t)}) K(\theta_j^{(t)}; \gamma_j, \theta_j^{(t)}/\gamma_j)} g(D_I, D_I^*)\right\}. $$

   That is, simulate $U \sim U(0, 1)$ and set $\theta_{1:2J}^{(t)} = \theta_j^*$ if $U < A\left(\theta_{1:2J}^{(t)}, \theta^*\right)$.

   v. If $100 \leq t \leq T_b$ and $\epsilon_t = \epsilon^{\text{min}}$ then check to see if require tuning of the proposal.

   Define the average acceptance probability over the last 100 iterations of updates for parameter $i$ by $\bar{a}_i^{(t-100:t)}$ and consider the adaption,

   $$\gamma_j^* = \begin{cases} 
   0.9\gamma_j & \text{if } \bar{a}_i^{(t-100:t)} < 0.3 \text{ and } \gamma_j > \gamma_j^{\text{min}}, \\
   1.1\gamma_j & \text{if } \bar{a}_i^{(t-100:t)} > 0.5, \\
   \gamma_j & \text{otherwise.} 
   \end{cases}$$

   Then set the proposal shape parameter as,

   $$\gamma_j = \max\{\gamma_j^*, \gamma_j^{\text{min}}\}.$$ 

The ABC-MCMC algorithm presented can be enhanced by utilizing an idea of Gramacy et al. [12] in an ABC setting. This involves a combination of tempering the tolerance $\{\epsilon_t\}_{t=1:T}$ and importance sampling corrections.
5. Example 1: Synthetic data

The tuning of the proposal distribution in this study is done for the simplest "base" distance metric, the weighted Euclidean distance. To study the effect of the distance metric in a comparative fashion we shall keep the proposal distribution unchanged.

The first example we present has a claims triangle of size $I = J = 9$. In this example we fix the true model parameters, denoted by $\mathbf{f} = (f_0, \ldots, f_{J-1})$ and $\mathbf{\sigma}^2 = (\sigma^2_0, \ldots, \sigma^2_{J-1})$ and given in Table 2, used to generate the synthetic data set.

5.1. Generation of synthetic data

To generate the synthetic observations for $\mathcal{D}_I$ we generate randomly the first column, denoted $\mathcal{B}_0$. Then conditional on this realization of $\mathcal{B}_0$ we make use of the model given in (2.1) to generate the remaining columns of $\mathcal{D}_I$, ensuring the model assumptions are satisfied. This requires setting $C_{i,0}$ sufficiently large, for appropriate choices of $\mathbf{f}$ and $\mathbf{\sigma}^2$, and then sample i.i.d. realizations of $\varepsilon_{i,j} \sim \mathcal{U}[-\sqrt{3}, \sqrt{3}]$ which are used to obtain $\mathcal{D}_I$, see the observations in Table 2.

5.2. Sensitivity analysis and convergence assessment

We perform a sensitivity analysis, studying the impact of the distance metric on the mixing of the Markov chain in the case of joint estimation of the chain ladder factors and the variance parameters.

The pre-tuned coefficient of variation of the Gamma proposal distribution for each parameter of the posterior was performed using the following settings; $T_b = 50,000$, $T = 200,000$, $\epsilon_{\min} = 0.1$ and initial values $\gamma_j = 1$ for all $j \in \{1, \ldots, 2J\}$. Additionally, the prior parameters for the chain ladder factors $F_j$ were set as $(\alpha, \beta) = (2, 1.2/2)$ and the parameters for the variance parameters $\Xi_j^{-2}$ were set as $(a, b) = (2, 1/2)$.

After tuning the proposal distributions during burnin and rounding the shape parameters we found that $\gamma_j = 10$ for all $j \in \{1, \ldots, 2J\}$ produced average acceptance probabilities for each parameter between 0.3 and 0.5.

Then keeping the proposal distribution constant and using a common data set $\mathcal{D}_I$ we ran 3 versions of the ABC-MCMC algorithm for 200,000 samples corresponding to:

1. Scaled Euclidean distance and joint estimation of posterior for $F, \Xi^2$
2. Mahalanobis distance (modified) and joint estimation of posterior for $F, \Xi^2$
3. Manhattan "City Block" distance and joint estimation of posterior for $F, \Xi^2$. 

20
We estimate the three convergence diagnostics we presented in Section 4.1. The results of this analysis are presented as a function of Markov chain iteration \( t \) post burnin of 50,000 samples.

**Autocorrelation Function:** In Figure 1 we demonstrate the estimated autocorrelation functions for the Markov chains of the random variables \( F_0 \) and \( \Xi_0^2 \). Since the posterior parameters in this model are independent it is suitable to analyze just the marginal parameters to get a reasonable estimate of the mixing behavior of the MCMC-ABC algorithm on all the posterior parameters. The results demonstrate the degree of serial correlation in the Markov chains generated for these parameters as a function of lag time \( \tau \). The higher the decay rate in the tail of the estimated ACF as a function of \( \tau \), the better the mixing of the MCMC algorithm.

Due to the independence properties of this model there is little difference between results obtained for Scaled Euclidean and Mahalanobis distances. As shown in Appendix A the estimate of the covariance matrix is diagonal on all but the right lower \( 2 \times 2 \) block. Hence, we recommend using the simple Scaled Euclidean distance metric as it provided the best trade-off between simplicity and mixing performance.

**Geweke Time Series Diagnostic:** Figure 2 presents results for the Geweke time series diagnostic. Again, we present the results for the random variables \( F_0 \) and \( \Xi_0^2 \). Note, we used the posterior mean as the sample function and a set of increasing values for \( \tilde{T} \) from \( T_b + 5,000 \) increasing in steps of 5,000 samples to \( T \). In each case we split the chain in each ”window” given by \( \{ \theta^{(t)}_i \}_{t=1:T_1} \) and \( \{ \theta^{(t)}_i \}_{t=T_*:T} \) according to recommendations from Geweke et al. [9]. We then calculate the convergence diagnostic \( Z_{\tilde{T}} \) which is the difference between these 2 means divided by the asymptotic standard error of their difference. As the chain length increases \( \tilde{T} \to \infty \), then the sampling distribution of \( Z \to N(0,1) \) if the chain has converged. Hence values of \( Z_{\tilde{T}} \) in the tails of a standard normal distribution suggest that the chain was not fully converged early on (i.e. during the 1st window). Hence, we plot \( Z_{\tilde{T}} \) scores versus increasing \( \tilde{T} \) and monitor if they lie within a 95% confidence interval \( Z_{\tilde{T}} \in [-1.96, 1.96] \).

The results in Figure 2 demonstrate that clearly the convergence properties of the distance functions differ. Again this is more material in the Markov chain for the variance parameter when compared to the Markov chain results for the chain ladder factor. The main point we note is that again one would advise against use of the ”City block” distance metric.

**Gelman and Rubin R statistic:** In Figure 3 we present the Gelman and Rubin convergence diagnostic. To calculate this we ran 20 chains in parallel, each of length 10,000 samples and for each chain we discarded 250 samples as burnin. We then estimated the \( R \) statistic as a function of simulation time post burnin. In Figure 3 we demonstrate the convergence rate of
the $R$ statistic to 1 for each distance metric on increasing blocks of 200 samples. Using this summary statistic all three distance metrics are very similar in terms of convergence rate of the $R$ statistic to 1.

Overall, these three convergence diagnostics demonstrate the simple scaled Euclidean distance metric is the superior choice. Secondly, we see appropriate convergence of the Markov chains under three convergence diagnostics which test different aspects of the mixing of the Markov chains, giving confidence in the performance of the MCMC-ABC algorithm in this model.

5.3. Parameter estimation Bayesian estimates

In this section we present results for the Scaled Euclidean distance metric, with a Markov chain of length 200,000 samples discarding the first 50,000 samples as burnin. In Tables 4 and ?? we present the Chain Ladder parameter estimates for the DFCL model and the associated parameter estimation error. We define the following quantities:

- $\hat{f}_j^{(MAP)}|_{\sigma_0;J-1}$, $\hat{f}_j^{(MMSE)}|_{\sigma_0;J-1}$, $\hat{\sigma}_{f_j}|_{\sigma_0;J-1}$ and $[\hat{q}_{0.05}, \hat{q}_{0.95}]|_{\sigma_0;J-1}$ denote respectively the Maximum a-Posteriori, Minimum Mean Square Error, posterior standard deviation of the conditional distribution of chain ladder factor $F_j$ and the posterior coverage probability estimates at 5% of the conditional distribution of chain ladder factor $F_j$, each of these estimates is conditional on knowledge of the true $\sigma_{0;J-1}$.

- $\hat{\sigma}_{f_j}^{(MAP)}$, $\hat{\sigma}_{f_j}^{(MMSE)}$, $\hat{\sigma}_{f_j}$ and $[\hat{q}_{0.05}, \hat{q}_{0.95}]$ denote the same quantities for the unconditional distribution after joint estimation of $F_{0;J-1}, \Xi_{0;J-1}$.

- $\text{Ave.}[A(\theta_1:2J, f_j)]$ and $\text{Ave.}[A(\theta_1:2J, \sigma_j)]$ denotes the average acceptance probability of the Markov chain.

- $\hat{\sigma}_{\sigma_j}^{2(MAP)}$, $\hat{\sigma}_{\sigma_j}^{2(MMSE)}$, $\hat{\sigma}_{\sigma_j}^2$ and $[\hat{q}_{0.05}, \hat{q}_{0.95}]$ denotes the same quantities for the chain ladder variances as those defined above for chain ladder factors.

For the frequentist approach we obtain the standard error in the estimates by using 1,000 bootstrap realizations of $\{D_I^{(s)}\}_{s=1:1000}$ to obtain $\{\hat{f}_{(s)}^{(CCL)}, \hat{\sigma}_{(s)}^{2(CCL)}\}_{s=1:1000}$. We use these bootstrap samples to calculate the standard deviation in the estimates of the parameters in the classical frequentist CL approach, present in brackets (.), next to their corresponding estimators. The standard errors in the Bayesian parameter estimates are obtained by blocking the Markov chain into 100 blocks of length 1,500 samples and estimating the posterior quantities on each block.
6. Example 2: Real Claims Reserving data

In this example we consider estimation using real claims reserving data from Wüthrich-Merz [22], see Table 4. This yearly loss data is turned into annual cumulative claims and divided by 10,000 for the analysis in this example. We use the analysis from the previous study to justify use of the joint MCMC-ABC simulation algorithm with a Scaled Euclidean distance metric.

We pre-tuned the coefficient of variation of the Gamma proposal distribution for each parameter of the posterior. This was performed using the following settings; \( T_b = 50,000, \bar{T} = 200,000, \epsilon^{\min} = 10^{-5} \) and initial values \( \gamma_j = 1 \) for all \( j \in \{1, \ldots, 2J\} \). Here we make a strict requirement of the tolerance level to ensure we have accurate results from our ABC approximation. Additionally, the prior parameters for the chain ladder factors \( F_j \) were set as \( (\alpha_j, \beta_j) = (1, f^{(CL)}_j) \) and the parameters for the variance \( \Xi^{-2}_j \) priors were set as \( (a_j, b_j) = (1, \sigma^{(CL)}_j) \).

The code for this problem was written in Matlab and it took approximately 10 min to simulate 200,000 samples from the MCMC-ABC algorithm, on a Intel Xeon 3.4GHz processor with 2Gb RAM.

After tuning the proposal distributions during burnin we obtained rounded shape parameters \( \gamma_{1:9} = [50, 100, 500, 500, 50, 000, 20, 000, 100, 000, 2, 000, 000, 3, 000, 000] \) provided average acceptance probabilities between 0.3 and 0.5.

In Figures 4 we present box-whisker plots of estimates of the distributions of the parameters \( F_{0:J-1}, \Xi_{0:J-1} \) obtained from the MCMC-ABC algorithm, post burnin. Figure 5 presents the Bayesian MCMC-ABC empirical distributions of the ultimate claims, \( C_{i,J} \) for \( i = 1, \ldots, I \). In Table 5 we present the predicted cumulative claims for each year along with the estimates for the chain ladder factors and chain ladder variances under both the classical approach and the Bayesian model. We see that with this fairly vague prior specified, we do indeed obtain convergence of the MCMC-ABC based Bayesian estimates \( \hat{f}^{(MMSE)}, \hat{\sigma}^{(MMSE)} \) to the classical estimates \( f^{(CL)}, \sigma^{(CL)} \).

In Figure 6 we present a study of the histogram estimate of the marginal posterior distribution for chain ladder factor \( \pi \left( f_0 | D_I, \epsilon^{\min} \right) \). The plot was obtained by sampling from the full posterior \( \pi \left( f, \sigma | D_I, \epsilon^{\min} \right) \) for each specified tolerance value, \( \epsilon^{\min} \). Then the samples for the particular chain ladder parameter in each plot are turned into a smoothed histogram estimate for each epsilon and plotted. We observe a significant difference in the parameter uncertainty as reflected by the change in the posterior precision as the tolerance decreases. Hence, the tolerance will impact the msep. Minimizing this impact, involves managing the computational budget to achieve mixing at a given tolerance level whilst minimizing the tolerance level.
Ultimately, we would like an algorithm which could work well for any $\epsilon^{\min}$, the smaller the better. However, we note that with a decreasing $\epsilon^{\min}$ in the sampler we present in this paper, one must take additional care to ensure the Markov chain is still mixing and not "stuck" in a particular state, as is observed to be the case in all MCMC-ABC algorithms. To avoid this acknowledged difficulty with MCMC-ABC requires that one either runs much longer MCMC chains or it requires the use of more sophisticated sampling algorithms such as SMC Samplers PRC-ABC based algorithms, see Sisson, Peters, Fan and Briers (2008). This is well beyond the scope of this paper and will be the subject of future papers and investigation in this context.

In Table 6 we present the predictive VaR at 95% and 99% levels for the ultimate predicted claims, obtained from the MCMC-ABC algorithm. These are easily obtained under the Bayesian setting, we simply used the MCMC-ABC posterior samples to explicitly obtain samples from the full predictive distribution of the cumulative claims after integrating out the parameter uncertainty numerically. In addition to this we present the analysis of the MSEP under the bootstrap frequentist procedure and the Bayesian MCMC-ABC and credibility estimates for the total predicted cumulative claims for each accident year $i$. We also present results for the sum of the total cumulative claims for each accident year. We can make the following conclusions from these results:

1. The process variance for each $C_{i,J}$ demonstrate that the unscaled conditional frequentist bootstrap and the credibility estimates are very close for all accident years $i$. The Bayesian results compare favorably.

2. The results for the parameter estimation error for the predicted cumulative claims $C_{i,J}$ demonstrate for small $i$ the Bayesian approach outperforms the frequentist approach. However, the Bayesian approach produces performance which worsens as $i$ increases, relative to the credibility approach. This could be a result of the tolerance level setting in the ABC algorithm. In future work it will be interesting to study the effect of the tolerance level on the parameter estimation error. Figure 6 demonstrates numerically that the tolerance $\epsilon$ should affect the parameter estimation error in the Bayesian approach. This study requires sophisticated algorithms such as those found in Sisson, Peters, Fan and Briers (2008).

3. The total results for the process variance for $C = \sum_i C_{i,J}$ demonstrate that the frequentist and credibility results are very close. Additionally, Bayesian total results are largest followed by credibility and then frequentist estimates which is in agreement with theoretical bounds.
4. The total results for the parameter estimation error for $C = \sum_i C_{i,J}$ demonstrate that frequentist unconditional bootstrap procedure results in the lowest total error. The Bayesian approach and credibility total parameter errors are close. Additionally, we note that the results in Wüthrich-Merz, (2008) Table 7.1 for the total parameter estimation error under an unconditional frequentist bootstrap with unscaled residuals is also very close to the total obtained for the frequentist approach.

7. Discussion

This paper has presented a distribution-free claims reserving model under a Bayesian paradigm. We have then present a novel advanced MCMC-ABC algorithm to obtain estimates from the resulting intractable posterior distribution of the chain ladder factors and chain ladder variances. We assessed several aspects of this algorithm, including the properties of the convergence of the MCMC algorithm as a function of the distance metric approximation in the ABC component. We studied the performance on a synthetic data set generated from known parameters, in order to demonstrate the accuracy of this methodology. Next we assessed a real claims reserving data set and compared the results we obtained for predicted cumulative ultimate claims to those obtained via classical chain ladder methods and via credibility theory. We note that this clearly demonstrates our algorithm is working accurately and provides us not only with the ability to obtain point estimates for the first and second moments of the ultimate cumulative claims, but an accurate empirical approximation of the entire distribution of the ultimate claims. This is valuable for many reasons, including prediction of reserves which are not based on centrality measures, such as the tail based VaR results we present. We have demonstrated a unique and accurate way in which one can bypass the need to make distributional approximations.

Acknowledgements

The first author thanks ETH FIM and ETH Risk Lab for their generous financial assistance whilst completing aspects of this work at ETH. The first author also thanks the Department of Mathematics and Statistics at the University of NSW for support through an Australian Postgraduate Award and to CSIRO for support through a postgraduate research top up scholarship. Finally, this material was based upon work partially supported by the National Science Foundation under Grant DMS-0635449 to the Statistical and Applied Mathematical Sciences Institute, North Carolina, USA. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National
References

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

[2] Del Moral, P., Doucet, A., Jasra, A. (2006). Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society Series. B.* 68(3), 411-436.

[3] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*. 7(1), 1-26.

[4] Efron, B., Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, NY.

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

[6] Fan, Y., Sisson, S.A., Peters, G.W. (2008). Improved efficiency in approximate Bayesian computation. *Technical report, Statistics Department, University of New South Wales*.

[7] Gelman, A., Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*. 4, 457-472.

[8] Gelman, A., Gilks, W.R., Roberts, G.O. (1997). Weak convergence and optimal scaling of random walk metropolis algorithm. *Annals of Applied Probability*. 7, 110-120.

[9] Geweke, J.F. (1991). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*. In J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (eds.) Bayesian Statistics, 4, Oxford University Press, Oxford.

[10] Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.

[11] Gisler, A., Wüthrich, M.V. (2008). Credibility for the chain ladder reserving method. *ASTIN Bulletin*. 38(2), 483-526.

[12] Gramacy, R.B., Samworth, R.J., King, R. (2008). Importance tempering. *Preprint, arXiv:0707.4242v5 [stat.Co]*.
[13] Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin* **23**, 213-225.

[14] Marjoram, P., Molitor, J., Plagnol, V., Tavare, S. (2003). Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Science USA* **100**, 15324-15328.

[15] Peters, G.W., Sisson, S.A. (2006). Bayesian inference, Monte Carlo sampling and operational risk. *Journal of Operational Risk* **1**(3), 27-50.

[16] Peters, G.W., Fan, Y., Sisson, S.A. (2008). On Sequential Monte Carlo, partial rejection control and approximate Bayesian computation. *Preprint, Statistics Department, University of New South Wales*.

[17] Peters, G.W., Sisson, S.A., Fan, Y. (2008). Design efficiency for "likelihood free" Sequential Monte Carlo samplers. *Preprint, Statistics Department, University of New South Wales*.

[18] Sisson, S.A., Peters, G.W., Fan, Y., Briers, M., (2008). Likelihood free samplers. *Preprint, Statistics Department, University of New South Wales*.

[19] Peters, G.W., Wüthrich, M.V., Shevchenko, P., (2008). Model risk in claims reserving within Tweedie's compound Poisson models. *to appear, ASTIN Bulletin*.

[20] Proakis, J.G., Manolakis, D.G., (1996). *Digital Signal Processing*. Upper Saddle River, N.J. Prentice Hall.

[21] Sisson, S.A., Fan, Y., Tanaka, M. (2007). Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Science USA* **104**, 1760-1765.

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

[23] Yao, J. (2008). Bayesian approach for prediction error in chain-ladder claims reserving. *Conference paper presented at the ASTIN Colloquium*, Manchester UK.
A. ABC algorithm

The ABC algorithm is typically justified in the simple rejection sampling framework. This then extends in a straightforward manner to other sampling frameworks such as the MCMC algorithm we utilise in this paper. We denote the posterior density from which we wish to draw samples by $\pi(\theta|y) \propto \pi(y|\theta) \pi(\theta)$ with $\theta \in \Omega$, where $\Omega$ denotes support of the posterior distribution and $\mathcal{Y}$ is the support for $y$.

The ABC method aims to draw from this posterior density $\pi(\theta|y)$ without the requirement of evaluating the computationally expensive or in our setting intractable likelihood $\pi(y|\theta)$. The cost of avoiding this calculation is that we obtain an "approximation".

1st case. We assume that the support $\mathcal{Y}$ is discrete. Given an observation $y \in \mathcal{Y}$ we would like to sample from $\pi(\theta|y)$. Then the original rejection sampling algorithm reads as follows:

**Rejection Sampling ABC**

1. Sample $\theta'$ from prior $\pi(\theta)$
2. Simulate synthetic data set of auxiliary variables $x|\theta' \sim \pi(x|\theta')$
3. ABC Rejection Condition: If $x = y$ then accept sample $\theta'$, else reject sample and return to step 1.

Then the chosen $\theta' \sim \pi(\theta|y)$ by simple rejection argument: Denote $\{x = y\}$ if $\theta'$ was chosen.

Then, the joint density of $(\theta', x)$, conditional on $\{y, x = y\}$, is given by

$$
\pi(\theta, x|y, x = y) = \frac{\pi(\theta, x|\theta, y)I\{y\}(x)}{\int \pi(\theta)|y\pi(y|\theta)d\theta} = \begin{cases} 
\frac{\pi(\theta,y)}{\pi(y)} = \pi(\theta|y) & \text{if } x = y, \\
0 & \text{otherwise.}
\end{cases}
$$

(A.1)

This implies that

$$
\sum_{x \in \mathcal{Y}} \pi(\theta, x|y, x = y) = \pi(\theta|y).
$$

(A.2)

Henceforth, this algorithm generates samples $\theta^{(t)} \sim \pi(\theta|y)$, for $t = 1, \ldots, T$.

2nd case. For more general supports $\mathcal{Y}$ one replaces the strict equality $x = y$ with a tolerance $\epsilon > 0$ and a measure of discrepancy or a distance metric, $\rho(x, y) \leq \epsilon$. In this case the posterior distribution is given by $\pi(\theta, x|y, \rho(x, y) < \epsilon)$. Implementing this algorithm in a rejection sampling framework gives the following:

**Rejection Sampling ABC**

1. Sample $\theta'$ from prior $\pi(\theta)$
2. Simulate synthetic data set of auxiliary variables $x|\theta' \sim \pi(x|\theta')$

3. ABC Rejection Condition 2: If $\rho(x, y) < \epsilon$ then accept sample $\theta'$, else reject sample and return to step 1.

In this case the joint density of $(\theta', x)$, conditional on $\{y, \rho(x, y) < \epsilon\}$, is given by

$$
\pi(\theta, x|y, \rho(x, y) < \epsilon) = \pi(\theta) \pi(x|\theta) I\{\rho(x, y) < \epsilon\}(x) dx d\theta.
$$

(A.3)

Note that for appropriate choices of the distance metric $\rho$ and assuming the necessary continuity properties for the densities we obtain that

$$
\lim_{\epsilon \to 0} \int \pi(\theta, x|y, \rho(x, y) < \epsilon) dx = \pi(\theta|y).
$$

(A.4)

This concept was taken further with the intention of improving the simulation efficiency by reducing the number of rejected samples. To achieve this sufficient statistics were used to replace the comparison between the auxiliary variables ("synthetic data") $x$ and the observations $y$. Denoting the sufficient statistics by $S(y)$ and $S(x)$, then this would allow one to decompose the likelihood under the Fisher-Neyman factorization theorem into, $\pi(y|\theta) = f(y)g(S(y)|\theta)$ for appropriate functions $f$ and $g$. In the ABC context presented above, the consequence of this decomposition is that when $\rho(S(y), S(x)) < \epsilon$ then we obtain samples from the posterior density $\pi(\theta, x|y, \rho(S(y), S(x)) < \epsilon$) similar to (A.3).

B. Scaling of statistics in distance metrics

In the Mahalanobis distance metric, estimation of the scaling weights given by the covariance $\Sigma_{D_I} = \text{Cov}(S(D^*_I; \tilde{\mu}, \tilde{s})|D_I)$, where $\tilde{\mu}$ and $\tilde{s}$ are the sample mean and standard deviation of $n$ i.i.d. residuals $\varepsilon_{i,j}$ (see also (4.4)-(4.5)). Next we outline the estimation of $\Sigma_{D_I}$ by a matrix $\hat{\Sigma}_{D_I}^{CL}$.

- Starting with the elements $\hat{\Sigma}_{D_I}^{CL}(k, l)$ with $k, l \in \{1, \ldots, n\}$, we obtain from the conditional resampling bootstrap

  $\text{Cov} \left(C^*_{i,j}, C^*_{i',j'}|D_I, \hat{f}^{(CL)}, \hat{\sigma}^{(CL)}\right) = 0$ if $i \neq i'$ or $j \neq j'$

  $\text{Var} \left(C^*_{i,j}|D_I, \hat{f}^{(CL)}, \hat{\sigma}^{(CL)}\right) = \hat{\sigma}^{2(CL)}_{j-1} C_{i,j-1}$.

- Considering the elements $k \in \{n+1, n+2\}, l \in \{1, \ldots, n\}$ and also $k \in \{1, \ldots, n\}, l \in \{n+1, n+2\}$ of the covariance matrix $\Sigma_{D_I}$, for simplicity we set $\hat{\Sigma}_{D_I}^{CL}(k, l) = 0$. 

29
• Considering elements $k, l \in \{n + 1, n + 2\}$, we assess now, $\text{Cov}(\tilde{\mu}, \tilde{s})$, either analytically or numerically by simulation of appropriate i.i.d. residuals.

**Parametric Approximation**

- In approximating $\tilde{\mu}$ and $\tilde{s}$ we assume i.i.d. samples $\varepsilon_{i,j} \sim \mathcal{N}(0, 1)$.

- Using the assumptions we know,
  
  \[
  \text{Var}(\tilde{\mu}) = \frac{1}{n}, \\
  \text{Var}(\tilde{s}) = \frac{1}{(n-1)n} \left[(1 + \frac{4}{n^2} + \frac{1}{n^2}) \sum_{s=1}^{n} \text{Var}(\tilde{\varepsilon}^2_s)\right] = \frac{1}{(n-1)n} \left[2n(1 + \frac{5}{n^2})\right], \\
  \text{Cov}(\tilde{\mu}, \tilde{s}) = \frac{1}{2(n-1)n^2} [1 - \frac{2}{n}].
  \]

- Under these assumptions,
  
  1. If the distribution of $\varepsilon_{i,j}$ is skewed then it is more appropriate to do a numerical approximation with the observed residuals from the bootstrap algorithm.
  2. The precision $\epsilon_t$ from the ABC-MCMC algorithm should depend on the size of the claims triangle, i.e. the number of residuals $n$.

**C. Estimating the Spectral Density**

This is calculated via a modified technique using Welch’s method, see Proakis-Manolakis [20], pages 910-913. This involves performing the following steps:

- Split each sequence, $\{\theta_i^{(t)}\}_{t=1:T_1}$ and $\{\theta_i^{(t)}\}_{t=T_1:T^*}$, into $L = 20$ non-overlapping blocks of length $N$.

- Apply a Hanning window function $w(t) = 0.5 \left(1 - \cos \left(\frac{2\pi t}{N-1}\right)\right)$ to the samples of the Markov chain in each block.

- Take the discrete Fourier transform (DFT) of each windowed block given by, $\tilde{\Theta}_i^{(t)}(k) = \sum_{t=0}^{N-1} \theta_i^{(t)} \exp \left(\frac{-2\pi ik}{N}\right)$.

- Estimate the spectral density (SD) as, $\hat{SD}(w_k) = \frac{1}{L} \sum_{t=0}^{L-1} \tilde{\Theta}_i^{(t)}(k)$. 

30
### Table 2: Synthetic Data - Cumulative claims $C_{i,j}$ for each accident year $i$ and development year $j$, $i + j \leq I$.

| Year | 0  | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  |
|------|----|----|----|----|----|----|----|----|----|----|
| 0    | 248.97 | 299.47 | 357.00 | 418.61 | 563.35 | 693.22 | 796.84 | 914.95 | 1,084.24 |
| 1    | 186.72 | 201.99 | 227.23 | 305.16 | 466.16 | 554.30 | 660.75 |
| 2    | 195.19 | 229.06 | 250.37 | 304.44 | 356.92 | 417.60 | 477.99 | 542.25 |
| 3    | 131.00 | 168.50 | 198.18 | 219.26 | 270.00 | 344.63 |
| 4    | 163.58 | 181.16 | 222.10 | 246.78 | 303.00 |
| 5    | 172.58 | 207.48 | 250.37 | 304.44 | 356.92 | 417.60 | 477.99 | 542.25 |
| 6    | 172.58 | 207.48 | 250.37 | 304.44 | 356.92 | 417.60 | 477.99 | 542.25 |
| 7    | 172.58 | 207.48 | 250.37 | 304.44 | 356.92 | 417.60 | 477.99 | 542.25 |
| 8    | 172.58 | 207.48 | 250.37 | 304.44 | 356.92 | 417.60 | 477.99 | 542.25 |
| 9    | 172.58 | 207.48 | 250.37 | 304.44 | 356.92 | 417.60 | 477.99 | 542.25 |

| $f_j$ | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 |
| $\sigma_j^2$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

### Table 3: Real Data - Incremental claims $Y_{i,j} = C_{i,j} - C_{i,j-1}$ for each accident year $i$ and development year $j$, $i + j \leq I$.

| Year | 0  | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  |
|------|----|----|----|----|----|----|----|----|----|----|
| 0    | 594.6975 | 372.1236 | 89.5717 | 20.7760 | 20.6704 | 6.2124 | 6.5813 | 1.4850 | 1.1130 | 1.5813 |
| 1    | 634.6756 | 324.6406 | 72.3222 | 15.1797 | 6.7824 | 3.6603 | 5.2752 | 1.1186 | 1.1646 |
| 2    | 626.9090 | 297.6223 | 84.7053 | 26.2768 | 15.2703 | 6.5444 | 5.3545 | 0.8924 |
| 3    | 586.3015 | 268.3224 | 72.2532 | 19.0653 | 13.2976 | 8.8340 | 4.3329 |
| 4    | 577.8885 | 274.5229 | 65.3894 | 27.3395 | 23.0288 | 10.5224 |
| 5    | 618.4793 | 282.8338 | 57.2765 | 24.8899 | 10.4957 |
| 6    | 560.0184 | 289.3207 | 56.3114 | 22.5517 |
| 7    | 528.8066 | 244.0103 | 52.8043 |
| 8    | 529.0793 | 235.7036 |
| 9    | 567.5508 |
Table 4: Comparison of Bayesian estimates for the chain ladder factors and variances versus Classical estimates, for synthetic data. Numerical standard errors in estimates are presented in brackets.
| Parameters | Year | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | $\hat{C}_{i,j}^{(CL)} - C_{i,j-1}$ |
|------------|-----|---|---|---|---|---|---|---|---|---|---|-------------------|
| $f^{(CL)}$ | 0   |   |   |   |   |   |   |   |   |   |   | 0                 |
| $f^{(MMSE)}$ | 0   |   |   |   |   |   |   |   |   |   |   | 0                 |
| $f^{(CL)}$ | 1   |   |   |   |   |   |   |   |   |   |   | 10,663,318        |
| $f^{(MMSE)}$ | 1   |   |   |   |   |   |   |   |   |   |   | 10,666,008        |
| $f^{(CL)}$ | 2   |   |   |   |   |   |   |   |   |   |   | 10,466,884        |
| $f^{(MMSE)}$ | 2   |   |   |   |   |   |   |   |   |   |   | 10,666,099        |
| $f^{(CL)}$ | 3   |   |   |   |   |   |   |   |   |   |   | 10,464,386        |
| $f^{(MMSE)}$ | 3   |   |   |   |   |   |   |   |   |   |   | 10,666,021        |
| $f^{(CL)}$ | 4   |   |   |   |   |   |   |   |   |   |   | 9,734,765          |
| $f^{(MMSE)}$ | 4   |   |   |   |   |   |   |   |   |   |   | 9,744,500          |
| $f^{(CL)}$ | 5   |   |   |   |   |   |   |   |   |   |   | 9,837,277          |
| $f^{(MMSE)}$ | 5   |   |   |   |   |   |   |   |   |   |   | 9,847,906          |
| $f^{(CL)}$ | 6   |   |   |   |   |   |   |   |   |   |   | 9,419,776          |
| $f^{(MMSE)}$ | 6   |   |   |   |   |   |   |   |   |   |   | 9,485,469          |
| $f^{(CL)}$ | 7   |   |   |   |   |   |   |   |   |   |   | 8,445,057          |
| $f^{(MMSE)}$ | 7   |   |   |   |   |   |   |   |   |   |   | 8,570,389          |
| $f^{(CL)}$ | 8   |   |   |   |   |   |   |   |   |   |   | 8,243,496          |
| $f^{(MMSE)}$ | 8   |   |   |   |   |   |   |   |   |   |   | 8,432,051          |
| $f^{(CL)}$ | 9   |   |   |   |   |   |   |   |   |   |   | 8,470,989          |
| $f^{(MMSE)}$ | 9   |   |   |   |   |   |   |   |   |   |   | 9,129,696          |
| $f^{(CL)}$ | 10  |   |   |   |   |   |   |   |   |   |   | 8,467,380          |
| $f^{(MMSE)}$ | 10  |   |   |   |   |   |   |   |   |   |   | 9,118,521          |

Table 5: Predicted cumulative CL claims $\hat{C}_{i,j}^{(CL)}$ for actual data and estimated CL reserves $\hat{C}_{i,j}^{(CL)} - C_{i,j-1}$ under the classical and Bayesian DFCL models.
| Accident Year i | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | Total |
|-----------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-------|
| \( C_{i,I-\hat{\Gamma}^{freq}_{i-I}}^{1/2} \) | 192 | 740 | 2,668 | 6,831 | 30,474 | 68,207 | 80,071 | 126,952 | 389,768 | 424,362 |
| \( C_{i,I-\Delta^{freq}_{i-I}}^{1/2} \) | 503 | 1,560 | 3,059 | 12,639 | 25,761 | 20,776 | 33,771 | 41,554 | 108,547 | 157,680 |
| \( \text{msep}_{C_{i,J}|D_I}^{freq} \left( \hat{C}_{i,J} \right) \)^{1/2} | 538 | 1,727 | 4,059 | 14,367 | 39,904 | 71,301 | 86,901 | 133,580 | 404,601 | 452,708 |
| Vco(\%) | 3.61% | 6.76% | 11.91% | 17.02% | 25.61% | 25.00% | 19.38% | 12.81% | 9.93% | 7.49% |
| | 134 | 533 | 2,307 | 7,185 | 27,367 | 74,235 | 86,404 | 129,038 | 437,482 | 470,982 |
| \( C_{i,I-\hat{\Gamma}^{Bayes}_{i-I}}^{1/2} \) | 224 | 894 | 1,801 | 4,327 | 15,819 | 29,861 | 49,188 | 152,879 | 211,633 |
| \( C_{i,I-\Delta^{Bayes}_{i-I}}^{1/2} \) | 261 | 1,040 | 2,927 | 8,387 | 31,610 | 80,016 | 92,224 | 138,099 | 463,425 | 504,934 |
| \( \text{msep}_{C_{i,J}|D_I}^{Bayes} \left( \hat{C}_{i,J} \right) \)^{1/2} | 354 | 2,183 | 5,632 | 15,820 | 61,122 | 152,531 | 173,665 | 816,701 | 910,757 |
| Vco(\%) | 1.75% | 4.07% | 8.59% | 10.06% | 21.86% | 13.68% | 11.86% | 8.53% | 910,757 |
| VaR_{0.95}^{Bayes} (C_{i,J} - E[C_{i,J}|D_I]|D_I) | 554 | 2,183 | 5,632 | 15,820 | 61,122 | 152,531 | 173,665 | 816,701 | 910,757 |
| VaR_{0.99}^{Bayes} (C_{i,J} - E[C_{i,J}|D_I]|D_I) | 726 | 2,918 | 7,430 | 22,515 | 79,472 | 201,322 | 228,448 | 211,125 | 462,941 |
| \( C_{i,I-\hat{\Gamma}^{cred}_{i-I}}^{1/2} \) | 192 | 740 | 2,668 | 6,831 | 30,474 | 68,207 | 80,071 | 126,952 | 389,768 | 424,362 |
| \( C_{i,I-\Delta^{cred}_{i-I}}^{1/2} \) | 188 | 534 | 1,493 | 3,391 | 13,515 | 27,284 | 29,674 | 43,901 | 129,764 | 185,015 |
| \( \text{msep}_{C_{i,J}|D_I}^{cred} \left( \hat{C}_{i,J} \right) \)^{1/2} | 269 | 913 | 3,057 | 7,627 | 33,337 | 73,462 | 85,392 | 134,329 | 462,941 |
| Vco(\%) | 1.81% | 3.58% | 8.97% | 9.04% | 21.40% | 25.77% | 19.04% | 12.88% | 10.40% | 7.82% |

Table 6: Comparison of the frequentist’s bootstrap \( msep^{freq} \), the Bayesian MCMC-ABC \( msep^{Bayes} \) and the credibility \( msep^{cred} \). The coefficient of variation is as defined in Wüthrich-Merz [22] see Equation
Estimated Autocorrelation Function (ACF) for parameters $F_0$ and $\Xi_0^2$.

Estimated Z scores for the posterior mean of parameters $F_0$ and $\Xi_0^2$ as a function of the length of the Markov chain $\tilde{T}$.

Estimated R statistic for parameters $F_0$ and $\Xi_0^2$ as a function of the length of the Markov chain $\tilde{T}$. 

35
Figure 4: Box-Whisker plots of parameters $F$ and $\Xi$ with each box marking the 25th, 50th, 75th percentiles. Top: 200,000 MCMC-ABC samples to estimate posterior for $F$. The sample mean and mode are denoted by ’*’ and ’o’ respectively. The classical estimators $\hat{f}^{(CL)}$ are denoted by $\triangle$. Bottom: 200,000 MCMC-ABC samples to estimate posterior for $\Xi$. The sample mean and mode are denoted by ’*’ and ’o’ respectively. The classical estimators $\hat{\sigma}^{(CL)}$ are denoted by $\triangle$.

Figure 5: Box-Whisker plots of predictive distribution of cumulative ultimate claims $C_{1:J}$ with the box marking the 25th, 50th, 75th percentiles, see also Table 6. The mean predicted ultimate claims under a Bayesian approach (using MMSE point estimates) are marked with *, the predicted mode for the ultimate claims (using MAP point estimates) is marked with o and the mean predicted ultimate claims under the DFCL classical method are marked with $\triangle$. 

36
Figure 6: Distribution of the chain ladder factor $F_0$ as a function of tolerance.