Chain Ladder Method: Bayesian Bootstrap versus Classical Bootstrap

Gareth W. Peters[1,2]   Mario V. Wüthrich[3]   Pavel V. Shevchenko[2]

[1] *UNSW Mathematics and Statistics Department, Sydney, 2052, Australia;*
email: peterga@maths.unsw.edu.au
[2] *CSIRO Mathematical and Information Sciences, Locked Bag 17, North Ryde, NSW, 1670, Australia*

[3] *ETH Zurich, Department of Mathematics, CH-8092 Zurich, Switzerland*

arXiv:1004.2548v1 [q-fin.CP] 15 Apr 2010

Chain Ladder Method: Bayesian Bootstrap versus Classical Bootstrap

Gareth W. Peters[1,2]   Mario V. Wüthrich[3]   Pavel V. Shevchenko[2]

Abstract

The intention of this paper is to estimate a Bayesian distribution-free chain ladder (DFCL) model using approximate Bayesian computation (ABC) methodology. We demonstrate how to estimate quantities of interest in claims reserving and compare the estimates to those obtained from classical and credibility approaches. In this context, a novel numerical procedure utilising Markov chain Monte Carlo (MCMC), ABC and a Bayesian bootstrap procedure was developed in a truly distribution-free setting. The ABC methodology arises because we work in a distribution-free setting in which we make no parametric assumptions, meaning we can not evaluate the likelihood point-wise or in this case simulate directly from the likelihood model. The use of a bootstrap procedure allows us to generate samples from the intractable likelihood without the requirement of distributional assumptions, this is crucial to the ABC framework. The developed methodology is used to obtain the empirical distribution of the DFCL model parameters and the predictive distribution of the outstanding loss liabilities conditional on the observed claims. We then estimate predictive Bayesian capital estimates, the Value at Risk (VaR) and the mean square error of prediction (MSEP). The latter is compared with the classical bootstrap and credibility methods.

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

1. Motivation

The distribution-free chain ladder model (DFCL) of Mack [14] is a 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 originally was 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 [30], Section 3.2). In the present paper we use the DFCL formulation to reproduce the CL reserves.

The paper presents a novel methodology for estimating a Bayesian DFCL model utilising a framework of approximate Bayesian computation (ABC) in a non-standard manner. A methodology utilising Markov chain Monte Carlo (MCMC), ABC and a Bayesian bootstrap procedure is developed in a distribution-free setting. The ABC framework is required because we work in a distribution-free setting in which we make no parametric assumptions about the form of the likelihood. Effectively, the ABC methodology allows us to overcome the fact that we cannot evaluate the likelihood point-wise in the DFCL model. Typically, ABC methodology circumvents likelihood evaluations by simulation from the likelihood. However, in this case simulation from the likelihood model is not directly available because no parametric assumption is made. We combine ABC methodology with bootstrap to overcome this additional complexity that the DFCL model presents in the ABC framework. Then, by using an MCMC numerical sampling algorithm combined with the novel version of ABC that has the embedded bootstrap procedure, we are able to obtain samples from the intractable posterior distribution of the DFCL model parameters.

This allows us to utilise this methodology to obtain the Bayesian posterior distribution of the DFCL model parameters empirically. Then we demonstrate two approaches in which we can utilise the posterior samples for the DFCL model parameters to obtain the Bayesian predictive distribution of the claims. The first approach involves using each posterior sample to numerically estimate the full predictive claims distribution given the observed claims. The alternative approach involves using the posterior samples for the DFCL model parameters to form Bayesian point estimators. Then, conditional on these point estimators, we can obtain the Bayesian conditional predictive distribution for the claims. The second approach will be relevant for

comparisons with the classical and credibility approaches. The first approach has the benefit that it integrates out of the Bayesian predictive claims distribution the parameter uncertainty associated with estimation of the DFCL chain ladder parameters.

The paper then 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 non-parametric bootstrap estimated prediction errors which can be obtained via one of two possible bootstrap procedures, conditional or unconditional. In this paper we consider the process of conditional back propagation; see [30] for in-depth discussion. These classical frequentist estimators are then compared to Bayesian point estimators. The Bayesian estimates considered are the maximum *a posteriori* (MAP) and the minimum mean square error (MMSE) estimators. For comparison with the classical frequentist reserve estimates, we also obtain the associated Bayesian estimated reserves conditional upon the Bayesian point estimators.

In addition, since in the Bayesian setting we obtain samples from the posterior for the parameters we use these along with the MSEP obtained by the estimated Bayesian point estimators to obtain associated posterior predictive intervals to be compared with the classical bootstrap procedures. We then robustify the prediction of reserves by Rao-Blackwellization, that is, we integrate out the influence of the unknown variance parameters in the DFCL model. Having done this, we analyse the resultant MSEP. This is again only achievable since in the Bayesian setting we obtain samples from the joint posterior for the CL factors and the variances.

To summarize our contribution, the novelty within this paper involves the development and comparison of a new estimation methodology to work with the Bayesian CL model for the DFCL model which makes no parametric assumptions on the form of the likelihood function; see also Gisler-Wüthrich [12]. This is unlike the works of Yao [31] and Peters et al. [21] 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 novel ABC methodology. In this regard we demonstrate that we do not need to make any parametric assumptions to perform posterior inference, avoiding potentially poor model assumptions made, as for example in the paper of Yao [31].

accident	development years j					
year i	0	1	$\ldots$	j	$\ldots$	I
0						
1		observed random variables $C_{i,j} \in \mathcal{D}_I$				
$\vdots$						
i						
$\vdots$		to be predicted $C_{i,j} \in \mathcal{D}_I^c$				
$I-1$						
I						

Table 1: Claims development triangles.

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 CL 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. This is followed by a discussion on classical versus Bayesian parameter estimators along with a bootstrap based procedure for the estimation of the 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 the 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, that is, $I = J$. At time

I, we have observations

$$\mathcal{D}_I = \{C_{i,j};\ i+j \le I\},\tag{2.1}$$

and for claims reserving at time I we need to predict the future claims

$$\mathcal{D}_I^c = \{C_{i,j};\ i+j > I,\ i \le I, j \le J\}.\tag{2.2}$$

Moreover, we define the set $\mathcal{B}_j = \{C_{i,k};\ i+k \le I, 0 \le k \le j\}$ for $j \in \{0,\ldots,I\}$, that is, $\mathcal{B}_0$ 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 is 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 $\mathcal{D}_I^c$. Set $\widehat{C}_{i,I-i} = C_{i,I-i}$ and for $j > I-i$

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

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 [14], the over-dispersed Poisson model (see England-Verrall [6]) or the Bayesian chain ladder model (see Gisler-Wüthrich [12]). 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 [12]).

Model Assumptions 2.1.

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

$$\pi(\boldsymbol{f},\boldsymbol{\sigma}) = \prod_{j=0}^{J-1} \pi(f_j)\,\pi(\sigma_j),\tag{2.4}$$

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

6

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

- Cumulative claims $C_{i,j}$ in different accident years i are independent.

- Cumulative claims satisfy the following time series representation

$$C_{i,j+1} = f_j C_{i,j} + \sigma_j \sqrt{C_{i,j}}\,\varepsilon_{i,j+1}, \tag{2.5}$$

where conditionally, given $\mathcal{B}_0$, we have that the residuals $\varepsilon_{i,j}$ are i.i.d. satisfying

$$E\left[\varepsilon_{i,j}|\mathcal{B}_0, \boldsymbol{F}, \boldsymbol{\Xi}\right] = 0 \quad \text{and} \quad \mathrm{Var}\left[\varepsilon_{i,j}|\mathcal{B}_0, \boldsymbol{F}, \boldsymbol{\Xi}\right] = 1, \tag{2.6}$$

and $P\left[C_{i,j} > 0 \middle| \mathcal{B}_0, \boldsymbol{F}, \boldsymbol{\Xi}\right] = 1$ 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-a.s.

Corollary 2.2. *Under Model Assumptions 2.1 we have that conditionally, given $\mathcal{D}_I$, the random variables $(F_0, \Xi_0), \dots, (F_{J-1}, \Xi_{J-1})$ are independent. Thus, we obtain the following posterior distribution for $(\boldsymbol{F}, \boldsymbol{\Xi})$, given $\mathcal{D}_I$,*

$$\pi\left(\boldsymbol{f}, \boldsymbol{\sigma}|\mathcal{D}_I\right) = \prod_{j=0}^{J-1} \pi\left(f_j, \sigma_j|\mathcal{D}_I\right). \tag{2.7}$$

This result follows from Theorem 3.2 in Gisler-Wüthrich [12]; from prior independence of the parameters; and the fact that $C_{i,j+1}$ only depends on F_j, Ξ_j and $C_{i,j}$ (Markov property). 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 $(\boldsymbol{F}, \boldsymbol{\Xi})$.

Model Assumptions 2.3.

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

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

- The variances $\Xi_j^2 \sim IG\left(a_j, b_j\right)$, where $IG\left(a_j, b_j\right)$ is an inverse gamma distribution with mean $E\left[\Xi_j^2\right] = b_j/(a_j - 1) = \widehat{\sigma}_j^{2(CL)}$ (see (3.1) below) and large variance.

Remarks

1. The likelihood model is intractable, meaning that no density can be written down analytically in the DFCL model. In formulating the Bayesian model we have only made distributional assumptions on the priors for the parameters $(\boldsymbol{F}, \boldsymbol{\Xi})$ but not on the observable cumulative claims $C_{i,j}$. Though we make distributional assumptions for the priors, the model is distribution-free because no distributional assumptions on the cumulative claims are made. As a result of only making assumptions on the priors, a standard Bayesian analysis using analytic posterior distributions cannot be performed. One way out of this dilemma would be to re-formulate the Bayesian model by making distributional assumptions (for example, this is done in Yao [31]) but then the model is no longer distribution-free. Another approach would be to use credibility methods (see Gisler-Wüthrich [12]) but this only gives statements for the first two moments. In the present set up we develop ABC methods that allow for a full distributional answer for the posterior distributions without making explicit distributional assumptions for the cumulative claims $C_{i,j}$.

2. Our priors are chosen as diffuse priors with large variances. This again highlights the differences between specification of the prior distributions and making distributional assumptions for the actual likelihood model, these are mutually exclusive ideas.

3. We select the priors to ensure that we maintain several relevant aspects of the DFCL model. In particular, 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 [31] failed in this aspect when it came to prior specification. Therefore we develop an alternative prior structure that satisfies these required properties of the DFCL model.

3. DFCL model parameter estimators

This section considers both classical and Bayesian estimators for the chain ladder framework, including both the chain ladder factors and the variance parameters.

3.1. Classical

In the classical CL method, the CL factors are estimated by $\widehat{f}_j^{(CL)}$ given in (2.3). The variance parameters are estimated by

$$\widehat{\sigma}_j^{2(CL)} \quad = \quad \frac{1}{I-j-1}\sum_{i=0}^{I-j-1} C_{i,j}\left(\frac{C_{i,j+1}}{C_{i,j}} - \widehat{f}_j^{(CL)}\right)^2; \tag{3.1}$$

see (3.4) in Wüthrich-Merz [30].

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 [14] for the last variance parameter which is defined by

$$\widehat{\sigma}_{J-1}^{2(CL)} = \min \left\{ \frac{\widehat{\sigma}_{J-2}^{4(CL)}}{\widehat{\sigma}_{J-3}^{2(CL)}}, \widehat{\sigma}_{J-3}^{2(CL)}, \widehat{\sigma}_{J-2}^{2(CL)} \right\}. \tag{3.2}$$

3.2. Bayesian

In a Bayesian inference context one calculates the posterior distribution of the parameters, given $\mathcal{D}_I$. As in (2.7) we denote this posterior by $\pi(\boldsymbol{f}, \boldsymbol{\sigma}|\mathcal{D}_I)$. Since the MCMC-ABC bootstrap procedure will allow us to obtain samples from the posterior distribution of the Bayesian DFCL model presented, we can now consider estimating CL point estimators using these samples.

There are two commonly used point estimators in Bayesian analysis that correspond to the posterior mode (MAP) and the posterior mean (MMSE), respectively:

$$\left(\widehat{f}_j^{(MAP)}, \widehat{\sigma}_j^{(MAP)} \right) = \arg\max_{f_j, \sigma_j} \pi(f_j, \sigma_j|\mathcal{D}_I), \tag{3.3}$$

and

$$\widehat{f}_j^{(MMSE)} = \int f_j \, \pi(f_j|\mathcal{D}_I) \, df_j = E[F_j|\mathcal{D}_I], \tag{3.4}$$

$$\widehat{\sigma}_j^{(MMSE)} = \int \sigma_j \, \pi(\sigma_j|\mathcal{D}_I) \, d\sigma_j = E[\Xi_j|\mathcal{D}_I]. \tag{3.5}$$

In the case in which f_j is not independent of σ_j, the MAP estimators obtained through joint maximization are optimal. However, in practice one often works with marginal estimators for simplicity. Additionally, note that for diffuse priors we find (see Corollary 5.1 in Gisler-Wüthrich [12])

$$\widehat{f}_j^{(MMSE)} \approx \widehat{f}_j^{(CL)}. \tag{3.6}$$

Hence, using Corollary 2.2, we obtain the approximation

$$
\begin{aligned}
E[C_{i,J}|\mathcal{D}_I] &= E[E[C_{i,J}|\mathcal{D}_I, \boldsymbol{F}, \boldsymbol{\Xi}]|\mathcal{D}_I] = C_{i,I-i} \, E\left[\prod_{j=I-i}^{J-1} F_j \middle| \mathcal{D}_I \right] \\
&= C_{i,I-i} \prod_{j=I-i}^{J-1} E[F_j|\mathcal{D}_I] = C_{i,I-i} \prod_{j=I-i}^{J-1} \widehat{f}_j^{(MMSE)} \\
&\approx C_{i,I-i} \prod_{j=I-i}^{J-1} \widehat{f}_j^{(CL)} = \widehat{C}_{i,J},
\end{aligned}
\tag{3.7}
$$

9

where on the last line we have an 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 [12].

3.3. Full predictive distribution and VaR

In addition, the posterior samples for the DFCL model parameters, obtained via the MCMC-ABC bootstrap procedure, will allow us to obtain the predictive distribution of the claims in two ways. The first is the full predictive distribution of the claims obtained after integrating out the posterior uncertainty associated with the Bayesian DFCL model parameters to empirically estimate

$$\pi\left(\mathcal{D}_I^c|\mathcal{D}_I\right) = \int \int \pi\left(\mathcal{D}_I^c|\boldsymbol{f},\boldsymbol{\sigma}\right)\pi\left(\boldsymbol{f},\boldsymbol{\sigma}|\mathcal{D}_I\right)d\boldsymbol{f}d\boldsymbol{\sigma}. \tag{3.8}$$

In practice, this numerical procedure involves taking each posterior sample for the DFCL model parameters and obtaining an estimate of the predicted claims.

The second approach involves using one of the Bayesian point estimators for the parameters such as the MMSE to obtain $\pi\left(\mathcal{D}_I^c|\widehat{\boldsymbol{f}}^{MMSE},\widehat{\boldsymbol{\sigma}}^{MMSE}\right)$. Alternatively, one may consider a Rao-Blackwellised version of the Bayesian predictive distribution of claims involving

$$\pi\left(\mathcal{D}_I^c|\widehat{\boldsymbol{f}}^{MMSE},\mathcal{D}_I\right) = \int \pi\left(\mathcal{D}_I^c|\widehat{\boldsymbol{f}}^{MMSE},\boldsymbol{\sigma}\right)\pi\left(\boldsymbol{\sigma}|\widehat{\boldsymbol{f}}^{MMSE},\mathcal{D}_I\right)d\boldsymbol{\sigma}$$

having numerically integrated out the Bayesian posterior uncertainty associated with the DFCL variance parameters. Such methods are typically known as empirical Bayesian approaches.

These results can then be applied to estimate any risk measures. For example, if we fix a security level 95% we can calculate the VaR on that level, which is defined by

$$\mathrm{VaR}_{0.95}\left(C_{i,J} - E\left[C_{i,J}|\mathcal{D}_I\right]\bigg|\mathcal{D}_I\right) = \min\left\{x;\ P\left[C_{i,J} - E\left[C_{i,J}|\mathcal{D}_I\right] > x\bigg|\mathcal{D}_I\right] \leq 0.05\right\}. \tag{3.9}$$

4. Bootstrap and mean square error of prediction

Assume that we have calculated the Bayesian predictor or the CL predictor given in (3.7). Then we would like to determine the prediction uncertainty, that is, 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; see (3.9).

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]. In the actuarial literature the development of bootstrap procedures includes the work of Taylor [27], Taylor-McGuire [28], [29], England-Verrall [5], [7] and Pinheiro et al. [19].

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 frequentist approach. That is, we assume that the CL factors $\boldsymbol{F} = \boldsymbol{f}$ and the standard deviation parameters $\boldsymbol{\Xi} = \boldsymbol{\sigma}$ given in Model Assumptions 2.1 are unknown constants. The bootstrap then generates synthetic data denoted by $\mathcal{D}_I^*$ that allow for the study of the fluctuations of $\widehat{\boldsymbol{f}}^{(CL)}$ and $\widehat{\boldsymbol{\sigma}}^{2(CL)}$ (for details see Section 7.4 in Wüthrich-Merz [30]). In the presented text we restrict ourselves to the conditional resampling approach presented in Section 7.4.2 of Wüthrich-Merz [30].

4.1. Non-parametric classical bootstrap (conditional version)

1. Calculate estimated residuals $\widetilde{\varepsilon}_{i,j}$ for $i + j \leq I$, $j > 0$, conditional on the estimators $\widehat{f}_{0:J-1}^{(CL)}$ and $\widehat{\sigma}_{0:J-1}^{2(CL)}$ and the observed data $\mathcal{D}_I$:

$$\widetilde{\varepsilon}_{i,j} \; = \; \widetilde{\varepsilon}_{i,j}(\widehat{f}_{j-1}^{(CL)}, \widehat{\sigma}_{j-1}^{(CL)}) \; = \; \frac{C_{i,j} - \widehat{f}_{j-1}^{(CL)} C_{i,j-1}}{\widehat{\sigma}_{j-1}^{(CL)} \sqrt{C_{i,j-1}}}.$$

2. These residuals $(\widetilde{\varepsilon}_{i,j})_{i+j \leq I}$ give the empirical bootstrap distribution $\widehat{F}_{\mathcal{D}_I}$.

3. Sample i.i.d. residuals $\widetilde{\varepsilon}_{i,j}^* \sim \widehat{F}_{\mathcal{D}_I}$ for $i + j \leq I$, $j > 0$.

4. Generate bootstrap observations (conditional resampling)

$$C_{i,j}^* = \widehat{f}_{j-1}^{(CL)} C_{i,j-1} + \widehat{\sigma}_{j-1}^{(CL)} \sqrt{C_{i,j-1}} \widetilde{\varepsilon}_{i,j}^*,$$

which defines $\mathcal{D}_I^* \; = \; \mathcal{D}_I^*(\widehat{\boldsymbol{f}}^{(CL)}, \widehat{\boldsymbol{\sigma}}^{(CL)})$. Note that for the unconditional version of bootstrap we should generate $C_{i,j}^* = \widehat{f}_{j-1}^{(CL)} C_{i,j-1}^* + \widehat{\sigma}_{j-1}^{(CL)} \sqrt{C_{i,j-1}^*} \widetilde{\varepsilon}_{i,j}^*$. For a discussion on this approach, see Section 7.4.1 of [30].

5. Calculate bootstrapped CL parameters $\widehat{f}_j^*$ and $\widehat{\sigma}_j^{2*}$ by

$$
\begin{aligned}
\widehat{f}_j^* &= \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}^*}{\sum_{i=0}^{I-j-1} C_{i,j}}, \\
\widehat{\sigma}_j^{2*} &= \frac{1}{I-j-1} \sum_{i=0}^{I-j-1} C_{i,j} \left(\frac{C_{i,j+1}^*}{C_{i,j}} - \widehat{f}_j^* \right)^2.
\end{aligned}
$$

6. Repeat steps 3-5 and obtain empirical distributions from the bootstrap samples $\widehat{C}_{i,J}^*$, $\widehat{f}_j^*$ and $\widehat{\sigma}_j^{2*}$. 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 $\widetilde{\varepsilon}_{i,j}$, in general, have variance smaller than 1 (see formula (7.23) in Wüthrich-Merz [30]). 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 [30] or England-Verrall [6]). Here, we use a different approach by introducing the novel Bayesian bootstrap method embedded within an MCMC-ABC algorithm to obtain empirically the posterior distribution of the Bayesian DFCL model, described below. Having obtained this, we can then calculate all required Bayesian parameter estimates, capital reserve estimates and associated risk measures such as VaR. Before presenting the methodology for this novel MCMC-ABC algorithm we will finalize this section with the decompositions of the MSEP under frequentist, Bayesian and credibility approaches.

4.2. Frequentist bootstrap estimates

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

$$
\begin{aligned}
\mathrm{msep}_{C_{i,J}|\mathcal{D}_I}\left(\widehat{C}_{i,J}\right) &= E\left[\left.\left(C_{i,J} - \widehat{C}_{i,J}\right)^2\right|\mathcal{D}_I\right] \qquad\qquad (4.1) \\
&= \mathrm{Var}\left(C_{i,J}|\mathcal{D}_I\right) + \left(E\left[C_{i,J}|\mathcal{D}_I\right] - \widehat{C}_{i,J}\right)^2.
\end{aligned}
$$

The first term is known as the conditional process variance and the second term as the parameter estimation uncertainty. In the frequentist approach (i.e. for given deterministic $\boldsymbol{F} = \boldsymbol{f}$ and $\boldsymbol{\Xi} = \boldsymbol{\sigma}$) these terms can be calculated as

$$\operatorname{Var}\left(C_{i,J} \middle| \mathcal{D}_I\right) = \left(E\left[C_{i,J} \middle| C_{i,I-i}\right]\right)^2 \sum_{j=I-i}^{J-1} \frac{\sigma_j^2/f_j^2}{E\left[C_{i,j} \middle| C_{i,I-i}\right]} \stackrel{def.}{=} C_{i,I-i}\Gamma_{I-i}, \qquad (4.2)$$

and

$$\left(E\left[C_{i,J} \middle| \mathcal{D}_I\right] - \widehat{C}_{i,J}\right)^2 = C_{i,I-i}^2 \left(\prod_{j=I-i}^{J-1} f_j - \prod_{j=I-i}^{J-1} \widehat{f}_j^{(CL)}\right)^2 \stackrel{def.}{=} C_{i,I-i}^2\Delta_{I-i}; \qquad (4.3)$$

see Wüthrich-Merz [30], Section 3.2.

The process variance (4.2) is estimated by replacing the parameters by its estimators,

$$\widehat{\operatorname{Var}}\left(C_{i,J} \middle| \mathcal{D}_I\right) = \left(\widehat{C}_{i,J}\right)^2 \sum_{j=I-i}^{J-1} \frac{\widehat{\sigma}_j^{2(CL)}/(\widehat{f}_j^{(CL)})^2}{\widehat{C}_{i,j}} \stackrel{def.}{=} C_{i,I-i}\widehat{\Gamma}_{I-i}^{freq}. \qquad (4.4)$$

The parameter estimation error is more involved and there we need the bootstrap algorithm. Assume that the bootstrap method gives T bootstrap samples $\widehat{f}_j^{*(1)}, \ldots, \widehat{f}_j^{*(T)}$. Then the parameter estimation error (4.3) is estimated by the sample variance of the product of the bootstrap observation chain ladder parameter estimates $\widehat{f}_j^{*(1)}, \ldots, \widehat{f}_j^{*(T)}$, which gives the estimator $C_{i,I-i}^2\widehat{\Delta}_{I-i}^{freq}$.

4.3. Bayesian estimates

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

$$\begin{aligned} \operatorname{msep}_{C_{i,J}|\mathcal{D}_I}\left(E\left[C_{i,J} \middle| \mathcal{D}_I\right]\right) &= \operatorname{Var}\left(C_{i,J} \middle| \mathcal{D}_I\right) \qquad\qquad (4.5) \\ &= E\left[\operatorname{Var}\left(C_{i,J} \middle| \mathcal{D}_I, \boldsymbol{F}, \boldsymbol{\Xi}\right) \middle| \mathcal{D}_I\right] + \operatorname{Var}\left(E\left[C_{i,J} \middle| \mathcal{D}_I, \boldsymbol{F}, \boldsymbol{\Xi}\right] \middle| \mathcal{D}_I\right). \end{aligned}$$

The average process variance is given by (see Wüthrich-Merz [30], Lemma 3.6)

$$E\left[\operatorname{Var}\left(C_{i,J} \middle| \mathcal{D}_I, \boldsymbol{F}, \boldsymbol{\Xi}\right) \middle| \mathcal{D}_I\right] = C_{i,I-i} \sum_{j=I-i}^{J-1} E\left[\prod_{m=I-i}^{j-1} F_m \, \Xi_j^2 \prod_{n=j+1}^{J-1} F_n^2 \,\middle|\, \mathcal{D}_I\right] \qquad (4.6)$$

$$= C_{i,I-i} \sum_{j=I-i}^{J-1} \prod_{m=I-i}^{j-1} E\left[F_m \middle| \mathcal{D}_I\right] E\left[\Xi_j^2 \middle| \mathcal{D}_I\right] \prod_{n=j+1}^{J-1} E\left[F_n^2 \middle| \mathcal{D}_I\right] \stackrel{def.}{=} C_{i,I-i}\widehat{\Gamma}_{I-i}^{Bayes},$$

where we have used posterior independence (2.7). The parameter estimation error is given by

$$\operatorname{Var}\left(E\left[C_{i,J} \middle| \mathcal{D}_I, \boldsymbol{F}, \boldsymbol{\Xi}\right] \middle| \mathcal{D}_I\right) = C_{i,I-i}^2 \operatorname{Var}\left(\prod_{j=I-i}^{J-1} F_j \,\middle|\, \mathcal{D}_I\right) \stackrel{def.}{=} C_{i,I-i}^2\widehat{\Delta}_{I-i}^{Bayes}, \qquad (4.7)$$

13

where we have used (3.7). Using (2.7), we obtain for the last term

$$C_{i,I-i}^2 \widehat{\Delta}_{I-i}^{Bayes} = C_{i,I-i}^2 \left[\prod_{j=I-i}^{J-1} E\left[F_j^2 \middle| \mathcal{D}_I \right] - \prod_{j=I-i}^{J-1} E\left[F_j \middle| \mathcal{D}_I \right]^2 \right]. \tag{4.8}$$

In order to calculate these two terms given in (4.6) and (4.8), we need to calculate the posterior distribution of $(\boldsymbol{F}, \boldsymbol{\Xi})$, given $\mathcal{D}_I$. 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 will be discussed in Section 5.

4.4. Credibility Estimates

As mentioned previously, we can also consider the credibility estimates given in Gisler-Wüthrich [12]. 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 [12]

$$\widehat{\mathrm{msep}}_{C_{i,J}|\mathcal{D}_I}\left(E\left[C_{i,J} \middle| \mathcal{D}_I \right] \right) = C_{i,I-i}\widehat{\Gamma}_{I-i}^{cred} + C_{i,I-i}^2 \widehat{\Delta}_{I-i}^{cred}, \tag{4.9}$$

where

$$\widehat{\Gamma}_{I-i}^{cred} = \sum_{j=I-i}^{J-1} \left\{ \prod_{m=I-i}^{j-1} \widehat{f}_m^{(CL)} \, \widehat{\sigma}_j^{2(CL)} \prod_{n=j+1}^{J-1} \left((\widehat{f}_n^{(CL)})^2 + \frac{\widehat{\sigma}_n^{2(CL)}}{\sum_{i=0}^{I-n-1} C_{i,n}} \right) \right\}, \tag{4.10}$$

$$\widehat{\Delta}_{I-i}^{cred} = \prod_{j=I-i}^{J-1} \left((\widehat{f}_j^{(CL)})^2 + \frac{\widehat{\sigma}_j^{2(CL)}}{\sum_{i=0}^{I-j-1} C_{i,j}} \right) - \prod_{j=I-i}^{J-1} (\widehat{f}_j^{(CL)})^2. \tag{4.11}$$

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).

5. ABC for intractable likelihoods and numerical Markov chain sampler

To estimate numerically the parameters, predicted claims and associated uncertainty measures such as the MSEP presented in the previous sections, the Bayesian approach requires the ability to sample from the posterior distribution of the DFCL model parameters. Obtaining samples $\left\{ \boldsymbol{f}^{(t)}, \boldsymbol{\sigma}^{2(t)} \right\}_{t=1:T}$ which are realisations of a random vector distributed with a posterior distribution $\pi\left(\boldsymbol{f}, \boldsymbol{\sigma} \middle| \mathcal{D}_I \right)$ in the DFCL model is difficult since the likelihood is intractable. Hence,

standard numerical approaches such as Markov chain Monte Carlo (MCMC) algorithms (see Gilks et al. [11]) cannot be directly used since they all require explicit repeated evaluation of the likelihood function at each stage of the Markov chain sampling algorithm. It is common to avoid this difficulty by making distributional assumptions for the form of the likelihood. This then violates the DFCL model assumption but allows for relatively standard sampling procedures to be applied. In this regard, one possible approach involves making a specific Gaussian assumption for the likelihood. One problem with this assumption, which is evident immediately, is that it precludes skewness in the model. Here, we do not make any such assumptions and instead we work in a truly distribution-free model using ABC to facilitate sampling from an intractable posterior distribution.

There is an additional complexity in the DFCL model not typically encountered when working with ABC methodology. Typically, ABC methodology is developed in the case in which the model likelihood cannot be evaluated point-wise, but conditional on parameter values, synthetic data is easily simulated from the model; see examples in Peters-Sisson [16] and Peters et al. [22]. This is not the case in the Bayesian DFCL model. Under the DFCL model the likelihood is only expressed by moment conditions, hence we cannot evaluate the likelihood point-wise and also the simulation from the likelihood cannot be performed directly. This is why we introduce the novel concept of the Bayesian bootstrap which is embedded within the ABC methodological framework.

Hence, 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 4.

ABC methods aim to sample from posterior distributions in the presence of computationally intractable likelihood functions. For an application in risk modelling of ABC methodology, see Peters-Sisson [16]. In this article we present a novel MCMC-ABC algorithm. Before presenting some details of the numerical MCMC procedure, we note that alternative numerical algorithms could be considered in the ABC context. For example, a sequential Monte Carlo (SMC) based algorithms which can improve simulation efficiency can be found in Del Moral et al. [2], Sisson et al. [25], Peters et al. [17],[18] and Marjoram et al. [15].

5.1. ABC methodology

In this section we provide a brief description of ABC methodology, which describes a suite of methods developed specifically for working with models in which the likelihood is computationally intractable. Here we work with a Bayesian model and consider the likelihood intractability to arise in the sense that we may not evaluate the likelihood point-wise.

The ABC method we consider here embeds an intractable target posterior distribution, in our case denoted by $\pi\left(\boldsymbol{f}, \boldsymbol{\sigma}|\mathcal{D}_I\right)$, into a general augmented model

$$\pi\left(\boldsymbol{f}, \boldsymbol{\sigma}, \mathcal{D}_I^*, \mathcal{D}_I\right) = \pi\left(\mathcal{D}_I|\mathcal{D}_I^*, \boldsymbol{f}, \boldsymbol{\sigma}\right) \pi\left(\mathcal{D}_I^*|\boldsymbol{f}, \boldsymbol{\sigma}\right) \pi\left(\boldsymbol{f}, \boldsymbol{\sigma}\right), \tag{5.1}$$

where $\mathcal{D}_I^*$ is an auxiliary vector on the same space as $\mathcal{D}_I$. In this augmented Bayesian model, the weighting function $\pi\left(\mathcal{D}_I|\mathcal{D}_I^*, \boldsymbol{f}, \boldsymbol{\sigma}\right)$ weights the intractable posterior. In this paper we consider the hierarchical model assumption, where we work with $\pi\left(\mathcal{D}_I|\mathcal{D}_I^*, \boldsymbol{f}, \boldsymbol{\sigma}\right) = g\left(\mathcal{D}_I|\mathcal{D}_I^*\right)$; see Reeves and Pettitt [24].

The mechanism in the ABC framework which allows one to avoid the evaluation of the intractable likelihood involves replacing this evaluation with data simulation from the likelihood. That is, given a realisation of the parameters of the model, a synthetic data set $\mathcal{D}_I^*$ is generated and compared to the original data set. This is a key aspect of the novel methodology we develop in this paper, since we utilise a bootstrap procedure to perform this simulation in the DFCL model setting.

Then summary statistics $S(\mathcal{D}_I^*)$ derived from this data are compared to summary statistics of the observed data $S(\mathcal{D}_I)$ and a distance $\rho\left(S(\mathcal{D}_I^*), S(\mathcal{D}_I)\right)$ is calculated. Finally, a weight is given to these parameters according to the weighting function $g\left(\mathcal{D}_I|\mathcal{D}_I^*\right)$, which may give greater weight when $S(\mathcal{D}_I^*)$ and $S(\mathcal{D}_I)$ are close (i.e. where $\rho\left(S(\mathcal{D}_I^*), S(\mathcal{D}_I)\right)$ is small).

For example, under the "Hard Decision" (HD) weighting given by

$$g\left(\mathcal{D}_I|\mathcal{D}_I^*\right) \propto \begin{cases} 1 & \text{if } \rho\left(S\left(\mathcal{D}_I\right), S\left(\mathcal{D}_I^*\right)\right) \leq \epsilon, \\ 0 & \text{otherwise;} \end{cases} \tag{5.2}$$

a reward is given to summary statistics of the augmented auxiliary variables $S\left(\mathcal{D}_I^*\right)$ within an ϵ-tolerance of the summary statistic of the actual observed data $S\left(\mathcal{D}_I\right)$, as measured by distance metric ρ.

Hence, in the ABC context, an approximation to the intractable target posterior marginal distribution $\pi\left(\boldsymbol{f}, \boldsymbol{\sigma}|\mathcal{D}_I\right)$, for which we are interested in formulating an empirical estimate, is

given by

$$\pi_{ABC}\left(\boldsymbol{f},\boldsymbol{\sigma}|\mathcal{D}_I,\epsilon\right) \propto \int g\left(\mathcal{D}_I|\mathcal{D}_I^*\right)\pi\left(\mathcal{D}_I^*|\boldsymbol{f},\boldsymbol{\sigma}\right)\pi\left(\boldsymbol{f},\boldsymbol{\sigma}\right)d\mathcal{D}_I^*. \tag{5.3}$$

As briefly mentioned, obtaining samples from the ABC posterior can be achieved using a number of numerical procedures, in this paper we consider an MCMC approach. The MCMC class of likelihood-free algorithm is justified on a joint space formulation, in which the stationary distribution of the Markov chain is given by $\pi_{ABC}\left(\boldsymbol{f},\boldsymbol{\sigma},\mathcal{D}_I^*|\mathcal{D}_I,\epsilon\right)$. The corresponding target distribution for the marginal distribution $\pi_{ABC}\left(\boldsymbol{f},\boldsymbol{\sigma}|\mathcal{D}_I,\epsilon\right)$ is then obtained via numerical integration. Note that the marginal posterior distribution $\pi_{ABC}\left(\boldsymbol{f},\boldsymbol{\sigma}|\mathcal{D}_I,\epsilon\right) \to \pi\left(\boldsymbol{f},\boldsymbol{\sigma}|\mathcal{D}_I\right)$ as $\epsilon \to 0$, recovering the "true" (intractable) posterior, assuming that $S\left(\mathcal{D}_I\right)$ are sufficient statistics and that the weighting function converges to a point mass on $S\left(\mathcal{D}_I\right)$ as $\epsilon \to 0$; see Peters-Sisson [16] and references therein for detailed discussion. Accordingly, the tolerance ϵ is typically set as low as possible for a given computational budget. In this paper we focus on the class of MCMC-based sampling algorithms.

The ABC methodology is novel both in the statistics literature and in the actuarial literature. It is informative to clearly provide the justification for this approach both theoretically and numerically. The simplest understanding of ABC is achieved by considering a rejection algorithm, therefore we provide a basic argument for how the ABC methodology works in simple rejection sampling in Appendix A. The actuarial DFCL model considered in this paper requires the more sophisticated MCMC-ABC methodology described below.

5.2. Technical justification for MCMC-ABC algorithm

For given observations $\mathcal{D}_I$ we want to sample from $\pi_{ABC}(\boldsymbol{f},\boldsymbol{\sigma}|\mathcal{D}_I)$ with an intractable likelihood function. We assume that $S(\mathcal{D}_I)$ is either the data itself or a summary of the data such as a sufficient statistic for the model from which we assume data $\mathcal{D}_I$ is a realisation. We assume that, given a set of parameters values $(\boldsymbol{f},\boldsymbol{\sigma})$, we can generate from the DFCL model (via a conditional bootstrap procedure) a synthetic data set denoted $\mathcal{D}_I^*$. We define a hard decision function $g(\mathcal{D}_I^*,\mathcal{D}_I) = \mathbb{I}\{\rho(S(\mathcal{D}_I^*),S(\mathcal{D}_I)) < \epsilon\}(\mathcal{D}_I^*)$ 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, (A.3)-(A.4), which gives us in the Bayesian DFCL model setting,

$$\pi_{ABC}(\boldsymbol{f},\boldsymbol{\sigma}|\mathcal{D}_I,\epsilon) = \frac{\int g(\mathcal{D}_I|\mathcal{D}_I^*)\,\pi(\mathcal{D}_I^*|\boldsymbol{f},\boldsymbol{\sigma})\,\pi(\boldsymbol{f},\boldsymbol{\sigma})\,d\mathcal{D}_I^*}{\int g(\mathcal{D}_I|\mathcal{D}_I^*)\,\pi(\mathcal{D}_I^*|\boldsymbol{f},\boldsymbol{\sigma})\,\pi(\boldsymbol{f},\boldsymbol{\sigma})\,d\mathcal{D}_I^*\,d\boldsymbol{f}\,d\boldsymbol{\sigma}} = \frac{\pi(\boldsymbol{f},\boldsymbol{\sigma})E\left[g(\mathcal{D}_I|\mathcal{D}_I^*)|\,\boldsymbol{f},\boldsymbol{\sigma}\right]}{E\left[g(\mathcal{D}_I|\mathcal{D}_I^*)\right]}.$$
$$\tag{5.4}$$

In the next step the numerator of (5.4) is approximated using the empirical distribution:

$$\pi(\boldsymbol{f},\boldsymbol{\sigma})E\left[g(\mathcal{D}_I|\mathcal{D}_I^*)|\,\boldsymbol{f},\boldsymbol{\sigma}\right] \approx \pi(\boldsymbol{f},\boldsymbol{\sigma})\frac{1}{L}\sum_{l=1}^{L}g\left(\mathcal{D}_I|\mathcal{D}_I^{*,(l)}(\boldsymbol{f},\boldsymbol{\sigma})\right), \qquad (5.5)$$

where $\mathcal{D}_I^{*,(l)}(\boldsymbol{f},\boldsymbol{\sigma}) \overset{i.i.d.}{\sim} \pi(\mathcal{D}_I^*|\boldsymbol{f},\boldsymbol{\sigma})$. Finally, we need to consider the denominator $E\left[g(X|y)\right]$. 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. Thus we use

$$
\begin{aligned}
\pi_{ABC}(\boldsymbol{f},\boldsymbol{\sigma}|\mathcal{D}_I,\epsilon) &\approx \frac{\int g(\mathcal{D}_I|\mathcal{D}_I^*)\,\pi(\mathcal{D}_I^*|\boldsymbol{f},\boldsymbol{\sigma})\,\pi(\boldsymbol{f},\boldsymbol{\sigma})\,d\mathcal{D}_I^*}{\int g(\mathcal{D}_I|\mathcal{D}_I^*)\,\pi(\mathcal{D}_I^*|\boldsymbol{f},\boldsymbol{\sigma})\,\pi(\boldsymbol{f},\boldsymbol{\sigma})\,d\mathcal{D}_I^*\,d\boldsymbol{f}\,d\boldsymbol{\sigma}} \\
&\propto \pi(\boldsymbol{f},\boldsymbol{\sigma})E\left[g(\mathcal{D}_I|\mathcal{D}_I^*)|\,\boldsymbol{f},\boldsymbol{\sigma}\right] \\
&\approx \pi(\boldsymbol{f})\pi(\boldsymbol{\sigma})\frac{1}{L}\sum_{l=1}^{L}g\left(\mathcal{D}_I|\mathcal{D}_I^{*,(l)}(\boldsymbol{f},\boldsymbol{\sigma})\right)
\end{aligned} \qquad (5.6)
$$

in order to obtain samples from $\pi_{ABC}(\boldsymbol{f},\boldsymbol{\sigma}|\mathcal{D}_I,\epsilon)$. 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 $g(\cdot|\cdot)$. The role of the distance measure ρ is evaluated by Peters et al. [22]. We further extend this analysis to the class of models considered in this paper. We analyse several choices for the distance measure ρ such as Mahlanobis distance, scaled Euclidean distance and the Manhattan "City Block" distance. Fan et al. [8] demonstrate that it is not efficient to utilise the standard Euclidean distance, especially when summary statistics considered are on different scales.

Additionally, using an MCMC-ABC algorithm, it is important to assess convergence diagnostics. Particularly when using MCMC-ABC 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 [10] time series statistic and the Gelman-Rubin [9] R-statistic convergence diagnostic in an ABC setting.

Concluding: We apply three different techniques in order to treat the intractable likelihood:

1. ABC is used to get a handle on the likelihood and therefore the intractable posterior.
2. As a result of using ABC we need to be able to generate synthetic data samples from the DFCL model given realisations of the parameters. These data samples come from the bootstrap algorithm.

18

3. We use a well understood MCMC based sampling algorithm that does not require calcula-
   tion of the non-analytic normalizing constants for the target distribution $\pi_{ABC}(\boldsymbol{f}, \boldsymbol{\sigma}|\mathcal{D}_I, \epsilon)$.
   The reason for this is that in the acceptance probability of the MCMC algorithm, the nor-
   malizing constant for the target posterior appears both in the numerator and denominator,
   resulting in cancellation.

The specific details of the MCMC algorithm and ABC choices are provided in the Appendix B.

6. Example 1: Analysis of MCMC-ABC bootstrap methodology on synthetic data

To test the accuracy of the methodology, first we use synthetic data generated with known
parameter values. 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 $\boldsymbol{f} = (f_0, \ldots, f_{J-1})$ and $\boldsymbol{\sigma}^2 = (\sigma_0^2, \ldots, \sigma_{J-1}^2)$ and given in
Table 2, used to generate the synthetic data set.

6.1. Generation of synthetic data

To generate the synthetic observations for $\mathcal{D}_I$, we generate randomly the first column (i.e. $\mathcal{B}_0$).
Then conditional on this realisation 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 $\boldsymbol{f}$ and $\boldsymbol{\sigma}^2$) and then sampling i.i.d. realisations
of $\varepsilon_{i,j} \sim \mathcal{U}\left[-\sqrt{3}, \sqrt{3}\right]$ used to obtain $\mathcal{D}_I$; see the observations in Table 2.

6.2. 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; $T_b = 50,000$, $\tilde{T} = 200,000$, $\epsilon^{\min} = 0.1$
and initial values $\gamma_j = 1$ for all $j \in \{1, \ldots, 2J\}$. Additionally, the prior parameters for the chain

19

ladder factors F_j were set as $(\alpha, \beta) = (2, 1.2/2)$ and the parameters for the variance parameters Ξ_j^{-2} were set as $(a, b) = (2, 1/2)$.

After tuning the proposal distributions during burn-in 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. This is a range typically used in practice when designing MCMC sampling algorithms.

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

1. scaled Euclidean distance and joint estimation of posterior for $\boldsymbol{F}, \boldsymbol{\Xi}^2$;
2. Mahlanobis distance (modified) and joint estimation of posterior for $\boldsymbol{F}, \boldsymbol{\Xi}^2$; and
3. Manhattan "City Block" distance and joint estimation of posterior for $\boldsymbol{F}, \boldsymbol{\Xi}^2$.

6.3. Convergence diagnostics

We estimate the three convergence diagnostics given in Appendix B. The results of this analysis are presented as a function of Markov chain iteration t post burn-in of 50,000 samples.

Autocorrelation Function: Figure 1 shows the estimated autocorrelation functions for the Markov chains of the random variables F_0 and Ξ_0^2. We analyze the marginal parameters to get a reasonable estimate of the mixing behavior of the MCMC-ABC algorithm. The results demonstrate the degree of serial correlation in the Markov chains generated for these parameters as a function of lag time τ. The higher the decay rate in the tail of the estimated ACF as a function of τ, 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 Mahlanobis distances. As shown in Appendix C, the estimate of the covariance matrix is diagonal on all but the right lower 2×2 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 shows results for the Geweke time series diagnostic. Again, we present the results for the random variables F_0 and Ξ_0^2. Note, we used the posterior mean as the sample function and a set of increasing values for $\widetilde{T}$ 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 $\{\theta_i^{(t)}\}_{t=1:T_1}$ and $\{\theta_i^{(t)}\}_{t=T^*:\widetilde{T}}$ according to recommendations from Geweke et al. [10]. We then

calculate the convergence diagnostic $Z_{\widetilde{T}}$ which is the difference between these two means divided by the asymptotic standard error of their difference. As the chain length increases $\widetilde{T} \to \infty$, the sampling distribution of $Z \to \mathcal{N}(0,1)$ if the chain has converged. Hence values of $Z_{\widetilde{T}}$ 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 $Z_{\widetilde{T}}$ scores versus increasing $\widetilde{T}$ and monitor if they lie within a 95% confidence interval $Z_{\widetilde{T}} \in [-1.96, 1.96]$. The results in Figure 2 clearly demonstrate 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: Figure 3 presents 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 burn-in. We then estimated the R statistic as a function of simulation time post burn-in. Figure 3 shows 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 that the simple scaled Euclidean distance metric is the superior choice. Secondly, we see appropriate convergence of the Markov chains under three convergence diagnostics which tests different aspects of the mixing of the Markov chains, giving confidence in the performance of the MCMC-ABC algorithm for this model.

6.4. Bayesian parameter 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 burn-in. Table 4 shows the CL parameter estimates for the DFCL model and the associated parameter estimation error. We define the following quantities:

- $\widehat{f}_j^{(MAP)}|\sigma_{0:J-1}$, $\widehat{f}_j^{(MMSE)}|\sigma_{0:J-1}$, $\widehat{\sigma}_{f_j}|\sigma_{0:J-1}$ 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 $\sigma_{0:J-1}$.

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

- $Ave[A(\theta_{1:2J}, f_j)]$ and $Ave[A(\theta_{1:2J}, \sigma_j)]$ denote the average acceptance probabilities of the Markov chain.

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

Note, the estimates for $\widehat{f}_j^{(MAP)}$ and $\widehat{\sigma}_j^{(MAP)}$ were obtained marginally. For the frequentist approach we obtain the standard error in the estimates by using 1,000 bootstrap realisations of $\left\{\mathcal{D}_I^{(s)}\right\}_{s=1:1,000}$ to obtain $\left\{\widetilde{\boldsymbol{f}}_{(s)}^{(CCL)}, \widetilde{\boldsymbol{\sigma}}_{(s)}^{2(CCL)}\right\}_{s=1:1,000}$. We use these bootstrap samples to calculate the standard deviation in the estimates of the parameters in the classical frequentist CL approach, given 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.

7. Example 2: Real Claims Reserving data

In this example we consider estimation using real claims reserving data from Wüthrich-Merz [30], see Table 3. 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: $T_b = 50,000$, $\widetilde{T} = 200,000$, $\epsilon^{\min} = 10^{-5}$ 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 $(\alpha_j, \beta_j) = \left(1, \widehat{f}_j^{(CL)}\right)$ and the parameters for the variance Ξ_j^{-2} priors were set as $(a_j, b_j) = \left(1, \widehat{\sigma}_j^{(CL)}\right)$. 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 Intel Xeon 3.4GHz processor with 2Gb RAM.

After tuning the proposal distributions during burn-in we obtained rounded shape parameters

$\gamma_{1:9} = [50; 100; 500; 500; 5,000; 20,000; 100,000; 2,000,000; 3,000,000]$ provided average acceptance probabilities between 0.3 and 0.5.

Estimates of f and σ

Figures 4 presents box-whisker plots of estimates of the distributions of the parameters $F_{0:J-1}$ and $\Xi_{0:J-1}$ obtained from the MCMC-ABC algorithm, post burn-in. Figure 5 shows 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 $\widehat{f}^{(MMSE)}, \widehat{\sigma}^{(MMSE)}$ to the classical estimates $\widehat{f}^{(CL)}, \widehat{\sigma}^{(CL)}$.

Dependence on tolerance ϵ

Figure 6 presents a study of the histogram estimate of the marginal posterior distribution for chain ladder factor $\pi_{ABC}\left(f_0 | \mathcal{D}_I, \epsilon^{\min}\right)$. The plot was obtained by sampling from the full posterior $\pi_{ABC}\left(\boldsymbol{f}, \boldsymbol{\sigma} | \mathcal{D}_I, \epsilon^{\min}\right)$ 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 $\epsilon^{\min}$ and plotted. The results of this analysis demonstrated that when ϵ is large, in this model greater than around $\epsilon^{\min} = 0.1$, the likelihood is not having an influence on the ABC posterior distribution. Hence, under an MCMC-ABC algorithm, this results in acceptance probabilities for the chain being artificially high, resulting in estimates of the posterior which reflect the prior distribution used (in this case a vague prior). As $\epsilon^{\min}$ is reduced, we notice that the changes in the estimate of the posterior distribution also reduces. The aim of this study is to demonstrate that once $\epsilon^{\min}$ reaches a small enough level, the effect of reducing it further is minimal on the posterior distribution. We see that changing $\epsilon^{\min}$ from 10^{-4} to 10^{-5} has not had a material impact on the posterior mean or variance, the change is less than 10%. As a result, reducing $\epsilon^{\min}$ past this point cannot be justified relative to the significant increase in computational effort required to achieve such a further reduction in $\epsilon^{\min}$.

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, one should run much longer MCMC chains or alternatively use of

more sophisticated sampling algorithms such as SMC Samplers PRC-ABC based algorithms; see Sisson et al. [25].

The conclusion of these findings is that a value of $\epsilon^{\min} = 10^{-5}$, which was used for the analysis of the data in this paper, is suitable numerically and computationally.

VaR and MSEP.

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, using 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, and the associated parameter uncertainty and process variance (see Section 4 for details).

We can make the following conclusions from these results:

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

2. The results for the parameter estimation error for the predicted cumulative claims $C_{i,J}$ demonstrate for small i that the Bayesian approach results in a smaller estimation error compared to the frequentist approach. For large i, the Bayesian approach produces larger estimation error relative to the credibility approach.

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 Table 7.1 of Wüthrich-Merz [30], for the total parameter estimation error under an unconditional frequentist bootstrap with unscaled residuals is also very close to the total obtained under the frequentist approach.

8. Discussion

This paper has presented a distribution-free claims reserving model under a Bayesian paradigm. A novel advanced MCMC-ABC algorithm was developed 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. The methodologies performance was demonstrated on a synthetic data set generated from known parameters. Next, it was applied to a real claims reserving data set. The results we obtained for predicted cumulative ultimate claims were compared to those obtained via classical chain ladder methods and via credibility theory. This clearly demonstrated that the 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 also with 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.

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. (1999). Analytic and bootstrap estimates of prediction errors in claims reserving. *Insurance: Mathematics and Economics.* **25**(3), 281-293.

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

[7] England, P.D., Verrall, R.J. (2007). Predictive distributions of outstanding liabilities in general insurance. *Annals of Actuarial Science.* **1**(2), 221-270.

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

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

[10] 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 Univeristy Press, Oxford.

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

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

[13] Gramacy, R.B., Samworth, R.J., King, R. (2008). Importance tempering. *Preprint, arXiv: 0707.4242v5 [stat.Co].*

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

[15] 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.

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

[17] 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.*

[18] 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.*

[19] Pinheiro, P.J.R., Andrade e Silva, J.M. and de Lourdes Centeno, M. (2003). Bootstrap methodology in claim reserving. *Journal of Risk Insurance* **70**(4), 701-714.

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

[21] Peters, G.W., Shevchenko, P., Wüthrich, M.V., (2008). Model uncertainty in claims reserving within Tweedie's compound Poisson models. *ASTIN Bulletin* **39**,1-33.

[22] Peters, G.W., Nevat, I., Sisson, S.A., Fan, Y., Yuan, J. (2009). Bayesian symbol detection in wireless relay networks. Accepted to appear in *IEEE Transactions on Signal Processing.*

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

[24] Reeves, R.W., Pettitt, A.N. (2005). A theoretical framework for approximate Bayesian computation. Presented at the International Workshop for Statistical Modelling, Sydney.

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

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

[27] Taylor, G.C. (1987). Regresion Models in Claims Analysis I: theory. *Proc. CAS* LXXIV, 354-383.

[28] Taylor, G.C., McGuire, G. (2005). Synchronous bootstrapping of seemingly unrelated regressions. Conference paper, *36th Astin Colloquium, 2005* Zurich, Switzerland.

[29] Taylor, G.C., McGuire, G. (2007). A synchronous bootstrap to account for dependencies between lines of business in the estimation of loss reserve prediction error. *North American Actuarial Journal* **8**(1), 37-44.

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

[31] Yao, J. (2008). Bayesian approach for prediction error in chain-ladder claims reserving. *Conference paper presented at the ASTIN Colloqium,* 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 Ω 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 θ' 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 θ', else reject sample and return to step 1.

Then the chosen θ' is distributed from $\pi(\theta|y)$. This follows from a simple rejection argument, Denote $\{x = y\}$ if θ' was chosen. Then, the joint density of (θ', x) conditional on $\{y, x = y\}$ is given by

$$\pi(\theta, x|y, x = y) = \frac{\pi(\theta)\pi(x|\theta)\mathbb{I}\{y\}(x)}{\int \pi(\theta)\pi(y|\theta)d\theta} = \begin{cases} \frac{\pi(\theta,y)}{\pi(y)} = \pi(\theta|y) & \text{if } x = y, \\ 0 & \text{otherwise.} \end{cases} \tag{A.1}$$

This implies that

$$\sum_{x \in \mathcal{Y}} \pi(\theta, x|y, x = y) = \pi(\theta|y). \tag{A.2}$$

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 poste-

rior 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 θ' 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 θ', else reject sample and return to step 1.

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

$$\pi(\theta, x|y, \rho(x, y) < \epsilon) = \frac{\pi(\theta) \ \pi(x|\theta) \ \mathbb{I}\{\rho(x, y) < \epsilon\}(x)}{\int \pi(\theta) \ \pi(x|\theta) \ \mathbb{I}\{\rho(x, y) < \epsilon\}(x) \ dx d\theta}. \tag{A.3}$$

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

$$\lim_{\epsilon \to 0} \int_{\mathcal{Y}} \pi(\theta, x|y, \rho(x, y) < \epsilon) dx = \pi(\theta|y). \tag{A.4}$$

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 $S(x)$, allows 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 $\rho(S(y), S(x)) < \epsilon$ the obtained samples are from the posterior density $\pi(\theta, x|y, \rho(S(y), S(x)) < \epsilon)$ similar to (A.3). In general, summary statistics will be used when sufficient statistics are not attainable.

B. MCMC-ABC to sample from $\pi_{ABC}(f, \sigma|\mathcal{D}_I)$

We develop an MCMC-ABC algorithm which has an adaptive proposal mechanism and annealing of the tolerance during burn-in of the Markov chain. Having reached the final tolerance post annealing, denoted $\epsilon^{\min}$, we utilise the remaining burn-in 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 Roberts et al. [20]. 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 it is the shape parameter of the Gamma proposal distribution. We impose an additional constraint that the minimum shape parameter value is set at $\gamma_j^{\min}$ for $j \in \{1, \ldots, 2J\}$.

MCMC-ABC algorithm using bootstrap samples.

1. For $t = 0$ initialize the parameter vector randomly, this gives $\theta_{1:2J}^{(0)} = \left(f_{0:J-1}^{(0)}, \sigma_{0:J-1}^{(0)} \right)$. Initialize the proposal shape parameters $\gamma_j \geq \gamma_j^{\min}$ 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 θ_j^* from a $\Gamma(\gamma_j, \theta_j^{(t)}/\gamma_j)$-distribution. We denote the Gamma proposal density by $K\left(\theta_j^*; \gamma_j, \theta_j^{(t)}/\gamma_j \right)$. This gives proposed parameter vector $\boldsymbol{\theta}^* = \left(\theta_{1:j-1}^{(t)}, \theta_j^*, \theta_{j+1:2J}^{(t)} \right)$.

      ii. Conditional on $\boldsymbol{\theta}^* = \left(\theta_{1:j-1}^{(t)}, \theta_j^*, \theta_{j+1:2J}^{(t)} \right)$, generate synthetic bootstrap data set $\mathcal{D}_I^* = \mathcal{D}_I^*(\boldsymbol{\theta}^*)$ using the bootstrap procedure detailed in Section 4 where we replace the CL parameter estimates $(\widehat{\boldsymbol{f}}^{(CL)}, \widehat{\boldsymbol{\sigma}}^{(CL)})$ by the parameters $\boldsymbol{\theta}^*$.

      iii. Evaluate summary statistics $S\left(\mathcal{D}_I; 0, 1\right)$ and $S\left(\mathcal{D}_I^*; \mu^*; s^*\right)$ and corresponding decision function $g(\mathcal{D}_I | \mathcal{D}_I^*)$ as described in Section 5.

      iv. Accept proposal with ABC acceptance probability

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

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

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

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

   Then set the proposal shape parameter as $\gamma_j = \max\{\gamma_j^*, \gamma_j^{\min}\}$.

The MCMC-ABC algorithm presented can be enhanced by utilising an idea of Gramacy et al. [13] in an ABC setting. This involves a combination of tempering the tolerance $\{\epsilon_t\}_{t=1:T}$ and importance sampling corrections.

B.1. ABC algorithmic choices for the time series DFCL model

We start with the choices of the ABC components.

- **Generation of a synthetic data set**: Note that in this setting not only is the likelihood intractable but also the generation of a synthetic data set $\mathcal{D}_I^*$ given the current parameter values $\boldsymbol{F}, \boldsymbol{\Xi}$ is not straightforward. The synthetic data set $\mathcal{D}_I^*$ is generated using the bootstrap procedure described in Section 4. Note that both the bootstrap residual $\widetilde{\varepsilon}_{i,j}$ and the bootstrap samples $\mathcal{D}_I^*$ are functions of the parameter choices; see Section 4.1. Therefore we generate for given $\boldsymbol{F} = \boldsymbol{f}$ and $\boldsymbol{\Xi} = \boldsymbol{\sigma}$ the bootstrap residuals $\widetilde{\varepsilon}_{i,j} = \widetilde{\varepsilon}_{i,j}(f_{j-1}, \sigma_{j-1})$ and the bootstrap samples $\mathcal{D}_I^* = \mathcal{D}_I^*(\boldsymbol{f}, \boldsymbol{\sigma})$ according to the non-parametric bootstrap (see Section 4.1) where we replace the CL parameter estimates $(\widehat{\boldsymbol{f}}^{(CL)}, \widehat{\boldsymbol{\sigma}}^{(CL)})$ by the parameters $\theta = (\boldsymbol{F}, \boldsymbol{\Xi})$.

- **Summary statistics**: We introduce summary statistics to replace sufficient statistics when they are not attainable for a given model. Then, in order to define the decision function g, we introduce summary statistics; see Appendix A. For the observed data $\mathcal{D}_I$ we define the vector

$$
\begin{aligned}
S\left(\mathcal{D}_I; 0, 1\right) &= (S_1, \ldots, S_{n+2}) \\
&= \left(C_{0,1}, \ldots, C_{0,J}, C_{1,1}, \ldots, C_{0,J-1}, \ldots, C_{I-2,1}, C_{I-2,2}, C_{I-1,1}; 0, 1\right),
\end{aligned}
$$

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

$$
\mu^* = \mu^*(\boldsymbol{F}, \boldsymbol{\Xi}) = \frac{1}{n} \sum_{i,j} \widetilde{\varepsilon}_{i,j}(F_{j-1}, \Xi_{j-1}), \tag{B.1}
$$

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

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

$$S\left(\mathcal{D}_I^*; \mu^*, s^*\right) = \left(C_{0,1}^*, \ldots, C_{0,J}^*, C_{1,1}^*, \ldots, C_{0,J-1}^*, \ldots, C_{I-2,1}^*, C_{I-2,2}^*, C_{I-1,1}^*; \mu^*, s^*\right).$$

- **Distance metrics**:

  - *Mahlanobis distance and scaled Euclidean distance*

    Here we draw on the analysis of Sisson et al. [8] that proposes the use of the Mahlanobis distance metric given by

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

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

    Note, the covariance matrix $\Sigma_{\mathcal{D}_I}$ 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

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

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

  $$g\left(\mathcal{D}_I | \mathcal{D}_I^*\right) = \mathbb{I}\left\{\rho\left(S\left(\mathcal{D}_I; 0, 1\right), S\left(\mathcal{D}_I^*; \mu^*, s^*\right)\right) < \epsilon\right\}.$$

- **Tolerance schedule**: We use the sequence

  $$\epsilon_t = \max\left\{20,000 - 10t, \epsilon^{\min}\right\}.$$

Note, the use of an MCMC-ABC 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

33

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 that can be directly used to determine, for a given computational budget, a suitable tolerance $\epsilon^{\min}$.

- **Convergence diagnostics:** We stress that when using an MCMC-ABC 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 burn-in period once tolerance of $\epsilon^{\min}$ is reached. If the total chain has length T, the initial burn-in stage will correspond to the first T_b samples and we define $\widetilde{T} = T - T_b$. We denote by $\{\theta_i^{(t)}\}_{t=1:\widetilde{T}}$ the Markov chain of the i-th parameter after burn-in. 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 $\{\theta_i^{(t)}\}_{t=1:\widetilde{T}}$, we define the biased autocorrelation estimate at lag τ by

    $$\widehat{ACF}(\theta_i, \tau) = \frac{1}{(\widetilde{T} - \tau)\hat{\sigma}(\theta_i)} \sum_{t=1}^{\widetilde{T}-\tau} [\theta_i^{(t)} - \widehat{\mu}(\theta_i)][\theta_i^{(t+\tau)} - \widehat{\mu}(\theta_i)], \qquad (B.3)$$

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

  - *Geweke [10] time series diagnostic.* For parameter θ_i it is calculated as follows:

    1. Split the Markov chain samples into two sequences, $\{\theta_i^{(t)}\}_{t=1:T_1}$ and $\{\theta_i^{(t)}\}_{t=T^*:\widetilde{T}}$, such that $T^* = \widetilde{T} - T_2 + 1$, and with ratios $T_1/\widetilde{T}$ and $T_2/\widetilde{T}$ fixed such that $(T_1 + T_2)/\widetilde{T} < 1$ for all $\widetilde{T}$.
    2. Evaluate $\widehat{\mu}\left(\theta_i^{T_1}\right)$ and $\widehat{\mu}\left(\theta_i^{T_2}\right)$ corresponding to the sample means on each sub sequence.
    3. Evaluate consistent spectral density estimates for each sub sequence, at frequency 0, denoted $\widehat{SD}(0; T_1, \theta_i)$ and $\widehat{SD}(0; T_2, \theta_i)$. The spectral density estimator considered in this paper is the classical non-parametric periodogram or power spectral

34

density estimator. We use Welch's method with a Hanning window; for details see Appendix D.

4. Evaluate convergence diagnostic given by
$$Z_{\widetilde{T}} = \frac{\widehat{\mu}\left(\theta_i^{T_1}\right) - \widehat{\mu}\left(\theta_i^{T_2}\right)}{T_1^{-1}\widehat{SD}(0;T_1,\theta_i) + T_2^{-1}\widehat{SD}(0;T_2,\theta_i)}.$$
According to the central limit theorem, as $\widetilde{T} \to \infty$ one has that $Z_{\widetilde{T}} \to \mathcal{N}(0,1)$ if the sequence $\{\theta_i^{(t)}\}_{t=1:\widetilde{T}}$ is stationary.

- *Gelman-Rubin [9] 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 five chains). For comparison purposes we split the total computational budget of $\widetilde{T}$ into $T_1 = T_2 = \ldots = T_5 = \frac{\widetilde{T}}{5}$. The convergence diagnostic for parameter θ_i is calculated using the following steps:

  1. Generate five independent Markov chain sequences, producing the chains for parameter θ_i denoted $\{\theta_{i,k}^{(t)}\}_{t=1:T_k}$ for $k \in \{1,\ldots,5\}$.
  2. Calculate the sample means $\widehat{\mu}\left(\theta_i^{T_k}\right)$ for each sequence and the overall mean $\widehat{\mu}\left(\theta_i^{\widetilde{T}}\right)$.
  3. Calculate the variance of the sequence means
     $\frac{1}{4}\sum_{k=1}^5 \left(\widehat{\mu}\left(\theta_i^{T_k}\right) - \widehat{\mu}\left(\theta_i^{\widetilde{T}}\right)\right)^2 \stackrel{def.}{=} B_i/T_k$.
  4. Calculate the within-sequence variances $\widehat{s}^2\left(\theta_i^{T_k}\right)$ for each sequence.
  5. Calculate the average within-sequence variance, $\frac{1}{5}\sum_{k=1}^5 \widehat{s}^2\left(\theta_i^{T_k}\right) \stackrel{def.}{=} W_i$.
  6. Estimate the target posterior variance for parameter θ_i by the weighted linear combination $\widehat{\sigma}^2\left(\theta_i^{\widetilde{T}}\right) = \frac{T_k-1}{T_k}W_i + \frac{1}{T_k}B_i$. 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 $\widetilde{T}$ but asymptotically, $\widetilde{T} \to \infty$, it converges to the posterior variance.
  7. Improve on the Gaussian estimate of the target posterior given by $\mathcal{N}(\widehat{\mu}\left(\theta_i^{\widetilde{T}}\right), \widehat{\sigma}^2\left(\theta_i^{\widetilde{T}}\right))$ 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 $\widehat{\mu}\left(\theta_i^{\widetilde{T}}\right)$, scale $\sqrt{\widehat{V}_i}$ and degrees of freedom df_i, each given respectively by:
     $\widehat{V}_i = \widehat{\sigma}^2\left(\theta_i^{\widetilde{T}}\right) + \frac{B_i}{\widetilde{T}}$ and $df_i = \frac{2\widehat{V}_i^2}{\widehat{Var}(\widehat{V}_i)}$, where the variance is estimated as

$$\widehat{\mathrm{Var}}\left(\widehat{V}_i\right) = \frac{1}{5}\left(\frac{T_1-1}{T_1}\right)^2\widehat{\mathrm{Var}}\left(\widehat{s}^2\left(\theta_i^{T_k}\right)\right) + \left(\frac{6}{\sqrt{2\widetilde{T}}}\right)^2 B_i^2$$

$$+ \frac{12(T_1-1)}{25T_1}\widehat{\mathrm{Cov}}\left(\widehat{s}^2\left(\theta_i^{T_k}\right),\widehat{\mu}\left(\theta_i^{\widetilde{T}}\right)\right) \qquad (\text{B.4})$$

$$- \frac{24(T_1-1)}{25T_1}\widehat{\mu}\left(\theta_i^{\widetilde{T}}\right)\widehat{\mathrm{Cov}}\left(\widehat{s}^2\left(\theta_i^{T_k}\right),\widehat{\mu}\left(\theta_i^{\widetilde{T}}\right)\right).$$

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 $\sqrt{\widehat{R}} = \sqrt{\frac{\widehat{V}_i df_i}{W_i(df_i-2)}}$, where as $\widetilde{T} \to \infty$ one can prove that $\widehat{R} \to 1$. This convergence diagnostic monitors the scale factor by which the current distribution for θ_i may be reduced if simulations are continued for $\widetilde{T} \to \infty$.

C. Scaling of statistics in distance metrics

In the Mahlanobis distance metric, estimation of the scaling weights is given by the covariance $\Sigma_{\mathcal{D}_I} = \mathrm{Cov}\left(S\left(\mathcal{D}_I^*;\widetilde{\mu},\widetilde{s}\right)|\mathcal{D}_I\right)$, where $\widetilde{\mu}$ and $\widetilde{s}$ are the sample mean and standard deviation of n i.i.d. residuals $\varepsilon_{i,j}$ (see also (B.1)-(B.2)). Next we outline the estimation of $\Sigma_{\mathcal{D}_I}$ by a matrix $\widehat{\Sigma}_{\mathcal{D}_I}^{CL}$.

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

  - $\mathrm{Cov}\left(C_{i,j}^*, C_{i',j'}^* \,\middle|\, \mathcal{D}_I, \widehat{\boldsymbol{f}}^{(CL)}, \widehat{\boldsymbol{\sigma}}^{(CL)}\right) = 0$ if $i \neq i'$ or $j \neq j'$
  - $\mathrm{Var}\left(C_{i,j}^* \,\middle|\, \mathcal{D}_I, \widehat{\boldsymbol{f}}^{(CL)}, \widehat{\boldsymbol{\sigma}}^{(CL)}\right) = \widehat{\sigma}_{j-1}^{2(CL)}C_{i,j-1}$.

- 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 $\Sigma_{\mathcal{D}_I}$, for simplicity we set $\widehat{\Sigma}_{\mathcal{D}_I}^{CL}(k,l) = 0$.

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

  Parametric Approximation

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

  - Using the assumptions we know that:

    $\mathrm{Var}(\widetilde{\mu}) = \frac{1}{n}$,

$$\operatorname{Var}(\widetilde{s}) = \frac{1}{(n-1)^2}\left[\left(1 + \frac{4}{n^2} + \frac{1}{n^2}\right)\sum_{s=1}^{n}\operatorname{Var}\left(\widetilde{\varepsilon}_s^2\right)\right] = \frac{1}{(n-1)^2}\left[2n\left(1 + \frac{5}{n^2}\right)\right],$$

$$\operatorname{Cov}(\widetilde{\mu}, \widetilde{s}) = \frac{1}{2(n-1)^2}\left[1 - \frac{2}{n}\right].$$

    − 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 ϵ_t from the MCMC-ABC algorithm should depend on the size of the

        claims triangle, that is, the number of residuals n.

D. Estimating the Spectral Density

This is calculated via a modified technique using Welch's method; see Proakis-Manolakis [23], 910-913 . This involves performing the following steps:

- Split each sequence $\{\theta_i^{(t)}\}_{t=1:T_1}$ and $\{\theta_i^{(t)}\}_{t=T^*:\widetilde{T}}$ 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 $\tilde{\Theta}_i^l(k) = \sum_{t=0}^{N-1}\theta_i^{(t)}\exp\left(-\frac{2\pi i k t}{N}\right)$.

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

Year	0	1	2	3	4	5	6	7	8	9
0	248.97	299.47	357.00	418.61	473.63	563.35	693.22	796.84	914.95	1,084.24
1	186.72	201.99	227.23	271.18	305.16	379.37	466.16	554.30	660.75	
2	172.58	207.48	250.37	304.44	356.92	417.60	477.99	542.25		
3	195.19	229.06	290.83	320.11	367.60	469.93	543.40			
4	131.00	168.50	198.18	219.26	270.00	344.63				
5	163.58	181.16	222.10	246.78	303.00					
6	294.30	373.08	477.16	566.20						
7	529.31	577.71	805.95							
8	249.00	321.83								
9	140.41									
f_j	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2
σ_j^2	1	1	1	1	1	1	1	1	1	1

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	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	24.4899	10.4957					
6	560.0184	289.3207	56.3114	22.5517						
7	528.8066	244.0103	52.8043							
8	529.0793	235.7936								
9	567.5568									

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, $i + j \leq I$.

DFCL model	$j=0$	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$	$j=8$
f_j	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20
$\widehat{f}_j^{(CL)}$	1.20 (2.40E-2)	1.22 (3.27E-2)	1.16 (2.46E-2)	1.17 (2.44E-2)	1.23 (2.63E-2)	1.19 (2.78E-2)	1.16 (2.59E-2)	1.17 (2.10E-2)	1.19 (2.51E-2)
$\widehat{f}_j^{(MAP)}|\sigma_{0:J-1}$	1.07 (0.02)	1.19 (0.02)	1.05 (0.02)	1.04 (0.02)	1.10 (0.02)	1.08 (0.02)	0.97 (0.02)	1.19 (0.03)	1.14 (0.04)
$\widehat{f}_j^{(MMSE)}|\sigma_{0:J-1}$	1.19 (1.34E-2)	1.21 (1.38E-2)	1.18 (1.27E-2)	1.19 (1.30E-2)	1.17 (1.37E-2)	1.18 (1.53E-2)	1.20 (1.60E-2)	1.18 (1.73E-2)	1.19 (2.35E-2)
$\widehat{\sigma}_{f_j}|\sigma_{0:J-1}$	0.23 (4.00E-3)	0.22 (3.1E-3)	0.20 (3.1E-3)	0.21 (3.2E-3)	0.22 (3.9E-3)	0.27 (1.01E-2)	0.35 (1.24E-2)	0.44 (1.41E-2)	0.70 (1.60E-2)
$[\hat{q}_{0.05},\hat{q}_{0.95}]|\sigma_{0:J-1}$	[0.75,1.50]	[0.77,1.50]	[0.76,1.41]	[0.75,1.44]	[0.82,1.51]	[0.78,1.52]	[0.65,1.60]	[0.46,1.79]	[0.25,2.50]
$\widehat{f}_j^{(MAP)}$	1.15 (0.02)	1.13 (0.02)	1.06 (0.02)	1.09 (0.02)	1.15 (0.02)	1.19 (0.02)	1.12 (0.03)	1.08 (0.03)	1.06 (0.04)
$\widehat{f}_j^{(MMSE)}$	1.19 (0.01)	1.18 (0.01)	1.17 (0.01)	1.18 (0.01)	1.16 (0.01)	1.20 (0.02)	1.18 (0.03)	1.16 (0.02)	1.20 (0.02)
$\widehat{\sigma}_{f_j}$	0.24 (5.1E-3)	0.24 (4.4E-3)	0.23 (5.0E-3)	0.26 (5.8E-3)	0.25 (5.6E-3)	0.25 (5.7E-3)	0.40 (0.01)	0.49 (0.02)	0.68 (0.02)
$[\hat{q}_{0.05},\hat{q}_{0.95}]$	[0.66,1.48]	[0.74,1.54]	[0.67,1.42]	[0.65,1.47]	[0.74,1.50]	[0.74,1.50]	[0.22,1.54]	[0.35,1.95]	[0.1,2.50]
$Ave[A\left(\theta_{1:2J},f_j\right)]$	0.21	0.21	0.19	0.22	0.25	0.21	0.22	0.20	0.24
σ_j^2	1	1	1	1	1	1	1	1	1
$\widehat{\sigma}_j^{2(CL)}$	1.02 (0.29)	0.75 (1.44)	0.51 (1.02)	0.49 (0.91)	0.71 (1.18)	0.72 (1.89)	0.25 (1.84)	0.31 (1.40)	0.25 (0.77)
$\widehat{\sigma}_j^{2(MAP)}$	0.58 (0.06)	0.96 (0.06)	0.54 (0.05)	0.78 (0.05)	0.78 (0.05)	0.81 (0.04)	0.61 (0.04)	0.79 (0.04)	0.56 (0.04)
$\widehat{\sigma}_j^{2(MMSE)}$	1.11 (0.03)	1.18 (0.03)	1.14 (0.04)	1.31 (0.03)	1.29 (0.03)	1.19 (0.02)	1.16 (0.03)	1.14 (0.03)	1.05 (0.02)
$\widehat{\sigma}_{\sigma_j}$	0.83 (0.02)	0.79 (0.02)	0.82 (0.02)	0.80 (0.02)	0.79 (0.02)	0.72 (0.02)	0.77 (0.02)	0.78 (0.02)	0.71 (0.02)
$[\hat{q}_{0.05},\hat{q}_{0.95}]$	[0.33,2.89]	[0.33,2.79]	[0.25,2.91]	[0.32,2.87]	[0.33,2.82]	[0.27,2.59]	[0.21,2.66]	[0.17,2.62]	[0.22,2.42]
$Ave[A\left(\theta_{1:2J},\sigma_j\right)]$	0.23	0.24	0.24	0.23	0.24	0.24	0.24	0.24	0.25

Table 4: Comparison of Bayesian estimates for the chain ladder factors and variances versus classical estimates, in the case of synthetic data. Numerical standard errors in estimates are presented in brackets.

Parameters	Year	0	1	2	3	4	5	6	7	8	9	$\widehat{C}_{i,J}^{(CL)} - C_{i,I-i}$
$f^{(CL)}$	0											0
$f^{(MMSE)}$												0
$f^{(CL)}$	1										10,663,318	15,126
$f^{(MMSE)}$											10,663,099	14,907
$f^{(CL)}$	2									10,646,884	10,662,008	26,257
$f^{(MMSE)}$										10,646,386	10,661,291	25,541
$f^{(CL)}$	3								9,734,574	9,744,764	9,758,606	34,538
$f^{(MMSE)}$									9,734,765	9,744,500	9,758,143	34,074
$f^{(CL)}$	4							9,837,277	9,847,906	9,858,214	9,872,218	85,302
$f^{(MMSE)}$								9,835,850	9,846,669	9,856,516	9,870,315	83,400
$f^{(CL)}$	5						10,005,044	10,056,528	10,067,393	10,077,931	10,092,247	156,494
$f^{(MMSE)}$							10,005,302	10,055,329	10,066,390	10,076,456	10,090,563	154,811
$f^{(CL)}$	6					9,419,776	9,485,469	9,534,279	9,544,580	9,554,571	9,568,143	286,121
$f^{(MMSE)}$						9,400,832	9,466,638	9,513,971	9,524,436	9,533,961	9,547,308	265,286
$f^{(CL)}$	7				8,445,057	8,570,389	8,630,159	8,674,568	8,683,940	8,693,030	8,705,378	449,167
$f^{(MMSE)}$					8,437,023	8,545,017	8,604,832	8,647,856	8,657,369	8,666,026	8,678,159	421,947
$f^{(CL)}$	8			8,243,496	8,432,051	8,557,190	8,616,868	8,661,208	8,670,566	8,679,642	8,691,971	1,043,242
$f^{(MMSE)}$				8,236,916	8,417,305	8,525,046	8,584,722	8,627,645	8,637,136	8,645,773	8,657,877	1,009,148
$f^{(CL)}$	9		8,470,989	9,129,696	9,338,521	9,477,113	9,543,206	9,592,313	9,602,676	9,612,728	9,626,383	3,950,814
$f^{(MMSE)}$			8,467,380	9,118,521	9,318,217	9,437,490	9,503,553	9,551,070	9,561,577	9,571,138	9,584,538	3,908,970
	$\widehat{f}_j^{(CL)}$	1.4925	1.0778	1.0229	1.0148	1.0070	1.0051	1.0011	1.0010	1.0014		6,047,061
	$\sigma_j^{(CL)}$	135.253	33.803	15.760	19.847	9.336	2.001	0.823	0.219	0.059		5,918,083
	$\widehat{f}_j^{(MMSE)}$	1.4919	1.0769	1.0219	1.0128	1.0070	1.0050	1.0011	1.0010	1.0014		
	$\sigma_j^{(MMSE)}$	154.221	33.000	16.770	22.397	8.300	2.166	0.720	0.158	0.041		

Table 5: Predicted cumulative CL claims $\widehat{C}_{i,j}^{(CL)}$ for actual data and estimated CL reserves $\widehat{C}_{i,J}^{(CL)} - C_{i,J-i}$ under the classical and Bayesian DFCL models.

Accident Year i	1	2	3	4	5	6	7	8	9	Total
$\left(C_{i,I-i}\widehat{\Gamma}_{I-i}^{freq}\right)^{1/2}$	192	740	2,668	6,831	30,474	68,207	80,071	126,952	389,768	424,361
$\left(C_{i,I-i}^{2}\widehat{\triangle}_{I-i}^{freq}\right)^{1/2}$	503	1,560	3,059	12,639	25,761	20,776	33,771	41,554	108,547	157,680
$\left(\mathrm{msep}_{C_{i,J}\mid\mathcal{D}_I}^{freq}\left(\widehat{C}_{i,J}\right)\right)^{1/2}$	538	1,727	4,059	14,367	39,904	71,301	86,901	133,580	404,601	452,708
$Vco_i(\%)$	3.61%	6.76%	11.91%	17.02%	25.61%	25.00%	19.38%	12.81%	9.93%	7.49%
$\left(C_{i,I-i}\widehat{\Gamma}_{I-i}^{Bayes}\right)^{1/2}$	134	533	2,307	7,185	27,367	74,235	86,404	129,038	437,482	470,982
$\left(C_{i,I-i}^{2}\widehat{\triangle}_{I-i}^{Bayes}\right)^{1/2}$	224	894	1,801	4,327	15,819	29,861	32,243	49,198	152,879	211,633
$\left(\mathrm{msep}_{C_{i,J}\mid\mathcal{D}_I}^{Bayes}\left(\widehat{C}_{i,J}\right)\right)^{1/2}$	261	1,040	2,927	8,387	31,610	80,016	92,224	138,099	463,425	504,934
$Vco_i(\%)$	1.75%	4.07%	8.59%	10.06%	20.42%	30.16%	21.86%	13.68%	11.86%	8.53%
$\mathrm{VaR}_{0.95}^{Bayes}\left(C_{i,J}-E[C_{i,J}\mid\mathcal{D}_I]\mid\mathcal{D}_I\right)$	554	2,183	5,632	15,820	61,122	152,531	173,665	161,619	816,701	910,757
$\mathrm{VaR}_{0.99}^{Bayes}\left(C_{i,J}-E[C_{i,J}\mid\mathcal{D}_I]\mid\mathcal{D}_I\right)$	726	2,918	7,430	22,515	79,472	201,322	228,448	211,125	1,278,665	1,454,966
$\left(C_{i,I-i}\widehat{\Gamma}_{I-i}^{cred}\right)^{1/2}$	192	740	2,668	6,831	30,474	68,207	80,071	126,952	389,769	424,362
$\left(C_{i,I-i}^{2}\widehat{\triangle}_{I-i}^{cred}\right)^{1/2}$	188	534	1,493	3,391	13,515	27,284	29,674	43,901	129,764	185,015
$\left(\mathrm{msep}_{C_{i,J}\mid\mathcal{D}_I}^{cred}\left(\widehat{C}_{i,J}\right)\right)^{1/2}$	269	913	3,057	7,627	33,337	73,462	85,392	134,329	410,802	462,941
$Vco_i(\%)$	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 [30].
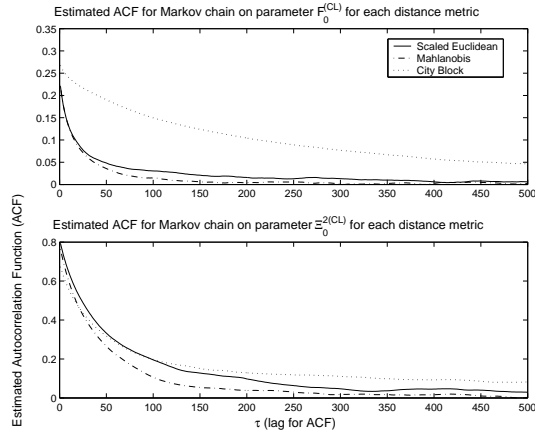
Figure 1: Estimated Autocorrelation Function (ACF) for parameters F_0 and Ξ_0^2.
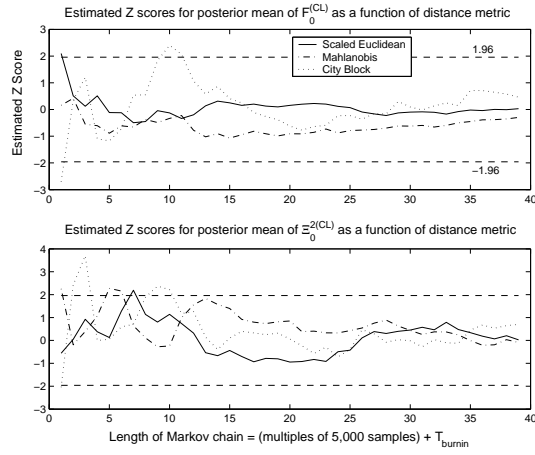

Figure 2: Estimated Z scores for the posterior mean of parameters F_0 and Ξ_0^2 as a function of the length of the Markov chain $\widetilde{T}$.
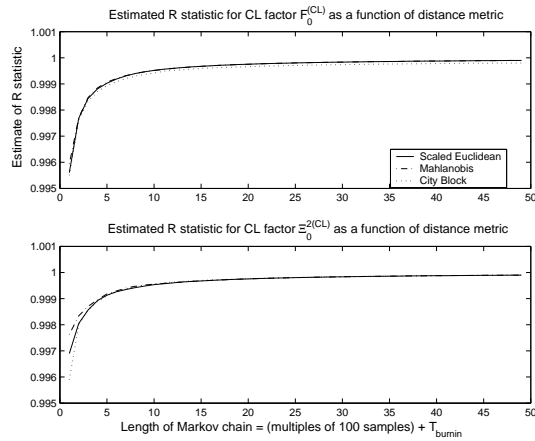

Figure 3: Estimated R statistic for parameters F_0 and Ξ_0^2 as a function of the length of the Markov chain $\widetilde{T}$.
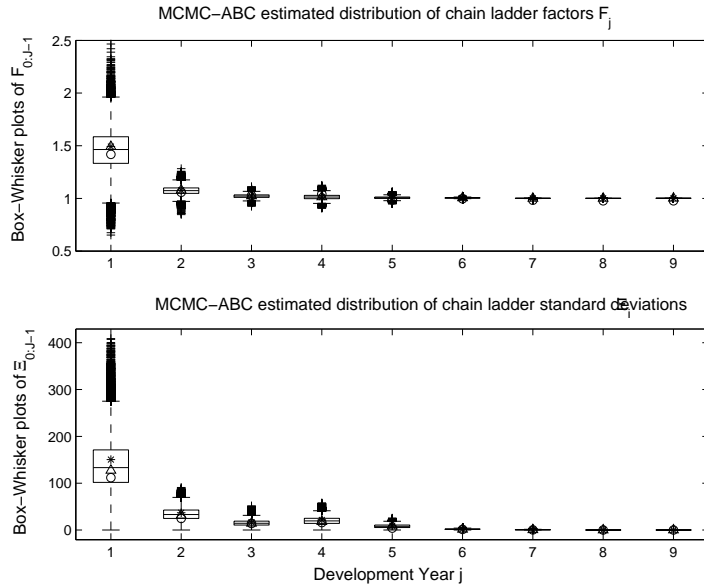
Figure 4: Box-Whisker plots of parameters $\boldsymbol{F}$ and $\boldsymbol{\Xi}$ with each box marking the $25^{th}, 50^{th}, 75^{th}$ percentiles. Top: 200,000 MCMC-ABC samples to estimate posterior for $\boldsymbol{F}$. The sample mean and mode are denoted by '*' and 'o' respectively. The classical estimators $\widehat{\boldsymbol{f}}^{(CL)}$ are denoted by $\triangle$. Bottom: 200,000 MCMC-ABC samples to estimate posterior for $\boldsymbol{\Xi}$. The sample mean and mode are denoted by '*' and 'o' respectively. The classical estimators $\widehat{\boldsymbol{\sigma}}^{(CL)}$ are denoted by '$\triangle$'.
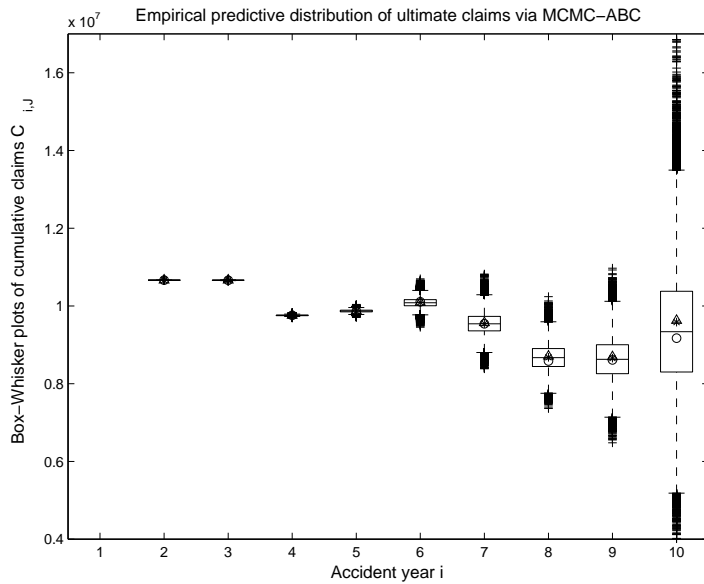

Figure 5: Box-Whisker plots of predictive distribution of cumulative ultimate claims $C_{1:J}$ with the box marking the $25^{th}, 50^{th}, 75^{th}$ percentiles; see also Table 6. The mean predicted ultimate claims under a Bayesian approach (using MMSE point estimates) are marked with '*', the predicted 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 '$\triangle$'.
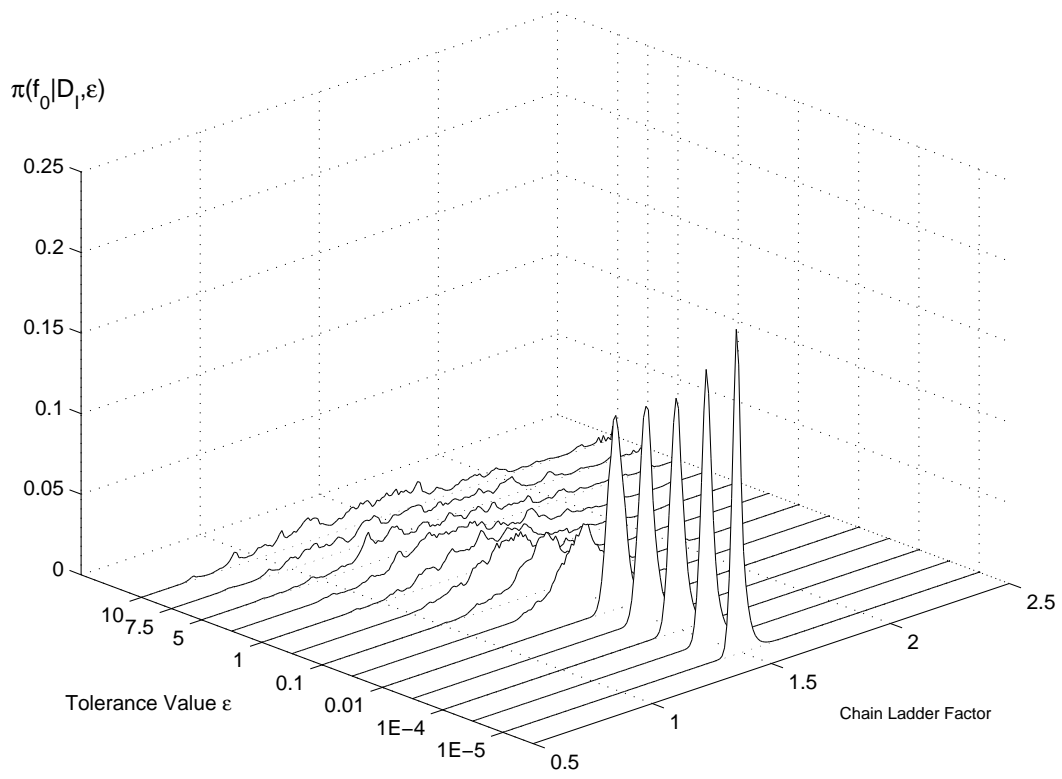
Figure 6: Distribution of the chain ladder factor F_0 as a function of tolerance.