1. Introduction

Bootstrapping has become very popular in stochastic claims reserving because of the simplicity and flexibility of the approach. One of the main reasons for this is the ease with which it can be implemented in a spreadsheet in order to obtain an approximation to the estimation error of a fitted model in a statistical context. Furthermore, it is also straightforward to extend it to obtain the approximation to the prediction error and the predictive distribution of a statistical process by including simulations from underlying distributions. Therefore, bootstrapping is a powerful tool for the most popular subject for reserving purposes in general insurance, the prediction error of the reserve estimates. It should be emphasised that to obtain the predictive distribution, rather than just the estimation error, it is necessary to extend the bootstrap procedure by simulating the process error. It is also important to realise that bootstrapping is not a "model", and therefore it is important to ensure that the underlying reserving models are correctly calibrated to the observed data. In this paper, we do not address the issue of model checking, but simply show how a bootstrapping procedure can be applied to the Munich chain ladder model.

In the area of non-life insurance reserving, there are primarily two types of data used. In addition to the paid claims triangle, there is frequently a triangle of incurred data also available. The traditional approach is to fit a model to either paid or incurred claims data, separately. One of the most popular methods used for reserving is the chain ladder technique. While we do not believe that this is the most appropriate approach for all data sets, it has retained its popularity for a number of reasons. For example, the parameters are understood in a practical context, it is flexible and it is easy to apply. This paper concentrates on methods which have a chain ladder structure, and in this context, two types of approaches exist: deterministic methods such as chain ladder, and the recently developed stochastic chain ladder reserving models. When the chain ladder technique is used (either as a deterministic approach or using a stochastic model), one set of data will be omitted - either the paid or the incurred data can be used, but not both at the same time. Obviously, this does not make full use of all the data available and results in the loss of some information contained in those data.

This leads us to consider whether it is possible to construct a model for both data sets, and to a consideration of the dependency between the two run-off triangles, which is not straightforward. This issue also arises when traditional methods are applied separately to each triangle, which produces inconsistent predicted ultimate losses. In response to this issue, Quarg and Mack (2004) proposed a different approach within a regression framework, considering the likely correlations between paid and incurred data. Quarg and Mack (2004) called this new method as the Munich chain ladder (MCL) model. It is this model that is the subject of this paper, and we show how the predictive distribution may be estimated using bootstrapping. Thus, in this paper an adapted bootstrap approach is described, combined with simulation for two dependent data sets. The spreadsheets used in this paper can be used in practice for any data sets, and are available on request from the authors.

The paper is set out as follows. Section 2 briefly describes the MCL model using a notation appropriate for this paper. In section 3, the basic algorithm and methodology of bootstrapping is explained. Section 4 shows how to obtain the estimates of the prediction errors and the empirical predictive distribution using the adapted bootstrapping and simulation methods. In Section 5, two numerical examples are provided including the data from Mack and also some real London market data. Finally, section 6 contains a discussion and conclusion.

2. The Munich chain ladder method

The MCL model aims to produce a more consistent ultimate loss prediction when modelling both paid and incurred claim data. It is specially designed to deal with the correlation between paid and incurred claims as the traditional models, such as chain ladder model, sometimes produce unsatisfactory results by ignoring this dependence. It should be emphasized that the paid and incurred claims from the same calendar years are not correlated. It is that the paid claims (incurred claims) are correlated to the incurred claims (paid claims) from the next (previous) calendar year.

The fundamental structure of the MCL model is the same as Mack's distribution-free chain ladder model. In the other word, the chain ladder development factors in the MCL model are obtained by Mack's distribution-free approach. More details of Mack's model are contained in Mack (1993).

Moreover, the MCL model adjusts the chain ladder development factors using the correlations between the observed paid and incurred claims. The adjusted chain ladder development factors therefore become individual not only for different development years but also for different accident years. The correlation adjustment is carried out within a linear modelling framework. This is explained in more detail in the sections 2.1 and 2.2.

2.1 Notation and Assumptions

For ease of notation, we assume that we have a triangle of data. Although the data could be classified in different ways, we refer to the rows as "accident years" and the columns as "development years". Denote C_{ij}^P as cumulative paid claims and C_{ij}^I as cumulative incurred claims occurred in accident year *i*, development year *j*, where $1 \le i \le n$ and $1 \le j \le n-i+1$ for the observed data. The aim of the chain ladder technique and of MCL is to estimate the data up to development year *n*. This produces estimates for $1 \le i \le n$ and $n-i+2 \le j \le n$, and we therefore refer to the complete rectangle of data in the assumptions: $1 \le i, j \le n$.

Mack's distribution-free chain ladder method models the pattern of the development factors in a regression framework, which are defined as $F_{ij}^P = \frac{C_{i,j+1}^P}{C_{ij}^P}$, for paid claims and $F_{ij}^I = \frac{C_{ij}^I}{C_{ij}^I}$, for incurred claims. Also the ratios of paid divided by incurred claims and the inverse are introduced as $Q_{ij} = \frac{C_{ij}^P}{C_{ij}^I}$ and $Q_{ij}^{-1} = \frac{C_{ij}^I}{C_{ij}^P}$, respectively.

Furthermore, define the observed data up to calendar year *k* as $P_k = \{C_{ij}^P : i+j-1 \le k\}$, $I_k = \{C_{ij}^I : i+j-1 \le k\}$ and $B_k = \{C_{i1}^P, C_{i1}^I : i+j-1 \le k\}$, for paid, incurred claims and both of these, respectively.

Assumptions A (Expectations)

(A1) For $1 \le j \le n$ there exists a constant f_j^P such that (for $i = 1, \ldots, n$)

$$E\left[F_{ij}^{P}\middle|P_{j-1}\right]=f_{j}^{P}.$$

This assumption is for paid claims. It is necessary to make another analogous assumption for incurred claims since both data sets are taken into account.

(A2) For $1\leq j\leq n$, there exists a constant f_{j}^{I} such that (for $i=1,...,n$)

$$E\left[F_{ij}^{I}\middle|I_{j-1}\right]=f_{j}^{I}.$$

In order to analyse the two run-off triangles dependently, the following assumptions are also required.

(A3) For $1\leq j\leq n$, there exists a constant q_{j}^{-1} such that (for $i=1,...,n$)

$$E\left[Q_{ij}^{-1}\middle|P_{j-1}\right]=q_{j}^{-1}.$$

(A4) For $1\leq j\leq n$, there exists a constant q_{j} such that (for $i=1,...,n$)

$$E\left[Q_{ij}\middle|I_{j-1}\right]=q_{j}.$$

Assumptions B (Variances)

(B1) For $1\leq j\leq n$, there exists a constant σ_{j}^{P} such that (for $i=1,...,n$)

$$Var\left[F_{ij}^{P}\middle|P_{j-1}\right]=\frac{\left(\sigma_{j}^{P}\right)^{2}}{C_{ij}^{P}}.$$

Again, the analogous assumption for the incurred claims is made as follows.

(B2) For $1\leq j\leq n$, there exists a constant σ_{j}^{I} such that (for $i=1,...,n$)

$$Var\left[F_{ij}^{I}\middle|I_{j-1}\right]=\frac{\left(\sigma_{j}^{I}\right)^{2}}{C_{ij}^{I}}.$$

Also, for the ratios of incurred to paid and vice versa, the following variance assumptions are made.

(B3) For $1\leq j\leq n$, there exists a constant τ_{j}^{P} such that (for $i=1,...,n$)

$$Var\left[Q_{ij}^{-1}\middle|P_{j-1}\right]=\frac{\left(\tau_{j}^{P}\right)^{2}}{C_{ij}^{P}}.$$

(B4) For $1\leq j\leq n$, there exists a constant τ_{j}^{I} such that for ($i=1,...,n$)

$$Var\left[Q_{ij}\left|I_{j-1}\right.\right]=\frac{\left(\tau_j^I\right)^2}{C_{ij}^I}.$$

Assumptions C (Independence)

(C1) The random variables pertaining to different accident years for paid claims, i.e. $\left\{C_{1j}^P \middle| j=1,2,...,n\right\}$, … , $\left\{C_{nj}^P \middle| j=1,2,...,n\right\}$, are stochastically independent.

(C2) The random variables pertaining to different accident years for incurred claims, i.e. $\left\{C_{1j}^I \middle| j=1,2,...,n\right\}$, … , $\left\{C_{nj}^I \middle| j=1,2,...,n\right\}$, are stochastically independent.

Using assumptions A to C, the Pearson residuals used in MCL model can be defined as shown in equations (4.1) to (4.4). These residuals are a crucial part of the bootstrapping procedures described in section 4.

$$r_{ij}^P=\frac{F_{ij}^P-E\left[F_{ij}^P\left|P_{j-1}\right.\right]}{\sqrt{Var\left[F_{ij}^P\left|P_{j-1}\right.\right]}}, \tag{2.1}$$

$$r_{ij}^{Q^{-1}}=\frac{Q_{ij}^{-1}-E\left[Q_{ij}^{-1}\left|P_{j-1}\right.\right]}{\sqrt{Var\left[Q_{ij}^{-1}\left|P_{j-1}\right.\right]}}, \tag{2.2}$$

$$r_{ij}^I=\frac{F_{ij}^I-E\left[F_{ij}^I\left|I_{j-1}\right.\right]}{\sqrt{Var\left[F_{ij}^I\left|I_{j-1}\right.\right]}}, \tag{2.3}$$

and

$$r_{ij}^Q=\frac{Q_{ij}-E\left[Q_{ij}\left|I_{j-1}\right.\right]}{\sqrt{Var\left[Q_{ij}\left|I_{j-1}\right.\right]}}. \tag{2.4}$$

Assumptions D (Correlations)

(D1) There exists a constant ρ^P such that (for $1 \le i,\, j \le n$)

$$E\left[r_{ij}^P\left|P_{j-1},I_{j-1}\right.\right]=\rho^P r_{ij}^{Q^{-1}}. \tag{2.5}$$

The following equation states that the constant ρ^P is in fact the correlation coefficient between the residuals. Note that since the residuals have variance 1, the correlation is equal to the covariance. The proof can be found in Quarg and Mack (2004).

$$Cov\left[r_{ij}^P,r_{ij}^{Q^{-1}}\left|P_{j-1},I_{j-1}\right.\right]=Corr\left[r_{ij}^P,r_{ij}^{Q^{-1}}\left|P_{j-1},I_{j-1}\right.\right]=\sqrt{\frac{Var\left[Q_{ij}^{-1}\left|P_{j-1}\right.\right]}{Var\left[F_{ij}^P\left|P_{j-1}\right.\right]}}Corr\left[F_{ij}^P,Q_{ij}^{-1}\left|P_{j-1},I_{j-1}\right.\right]=\rho^P \tag{2.6}$$

Quarg and Mack (2004) uses equation *(4.5)* to derive expected MCL paid development factors adjusted by the correlation as shown in equation *(4.6)*.

$$E\left[F_{ij}^{P}\left|P_{j-1},I_{j-1}\right.\right]=E\left[F_{ij}^{P}\left|P_{j-1}\right.\right]+Corr\left[F_{ij}^{P},Q_{ij}^{-1}\left|P_{j-1},I_{j-1}\right.\right]\left(Q_{ij}^{-1}-E\left[Q_{ij}^{-1}\left|P_{j-1}\right.\right]\right). \qquad (2.7)$$

(D2) Analogously to assumption *(D1)*, for the incurred claims it is assumed that there exists a constant ρ^{I} such that (for $1\le i,\, j\le n$)

$$E\left[r_{ij}^{I}\left|P_{j-1},I_{j-1}\right.\right]=\rho^{I}r_{ij}^{Q}. \qquad (2.8)$$

Similarly, the constant ρ^{I} measures the correlation coefficient or the covariance between the residuals. i.e.

$$Cov\left[r_{ij}^{I},r_{ij}^{Q}\left|C_{i,j-1}^{P},C_{i,j-1}^{I}\right.\right]=Corr\left[r_{ij}^{I},r_{ij}^{Q}\left|P_{j-1},I_{j-1}\right.\right]=\sqrt{\frac{Var\left[Q_{ij}\left|I_{j-1}\right.\right]}{Var\left[F_{ij}^{I}\left|I_{j-1}\right.\right]}}Corr\left[F_{ij}^{I},Q_{ij}\left|P_{j-1},I_{j-1}\right.\right]=\rho^{I} \qquad (2.9)$$

Hence, the expected MCL incurred development factors adjusted by the correlation can be derived from equation *(4.7)* as follows,

$$E\left[F_{ij}^{I}\left|C_{i,j-1}^{P},C_{i,j-1}^{I}\right.\right]=E\left[F_{ij}^{I}\left|C_{ij}^{I}\right.\right]+Cov\left[F_{ij}^{I},Q\ \left|P_{j-1},I_{j-1}\right.\right]\left(Q_{ij}-E\left[Q_{ij}\left|I_{j-1}\right.\right]\right). \qquad (2.10)$$

2.2 Unbiased Estimates of the Parameters

In this section, we summarise the unbiased estimates of the parameters derived by Quarg and Mack (2004). For the paid data, estimates are required for the parameters of the development factors, the variance constants and also the correlation coefficient.

The estimates of the paid development factor parameters can be interpreted as weighted averages of the observed development factors F_{ij}^{P} or Q_{ij}^{-1} , using C_{ij}^{P} as the weights:

$$\hat{f}_{j}^{P}=\frac{\sum_{i=1}^{n-j}C_{i,j+1}^{P}}{\sum_{i=1}^{n-j}C_{ij}^{P}}=\sum_{i=1}^{n-j}\frac{C_{ij}^{P}}{\sum_{i=1}^{n-j}C_{ij}^{P}}F_{ij}^{P} \qquad (2.11)$$

and

$$\hat{q}_j^{-1} = \frac{\sum_{i=1}^{n-j} C_{ij}^I}{\sum_{i=1}^{n-j} C_{ij}^P} = \sum_{i=1}^{n-j} \frac{C_{ij}^P}{\sum_{i=1}^{n-j} C_{ij}^P} Q_{ij}^{-1}. \tag{2.12}$$

The unbiased estimates of the variance constants coefficient are as follows:

$$\left(\hat{\sigma}_j^P\right)^2 = \frac{1}{n-j-1} \sum_{i=1}^{n-j} C_{ij}^P \left(F_{ij}^P - \hat{f}_j^P\right)^2 \tag{2.13}$$

and

$$\left(\hat{\tau}_j^P\right)^2 = \frac{1}{n-j} \sum_{i=1}^{n-j+1} C_{ij}^P \left(Q_{ij}^P - \hat{q}_j^{-1}\right)^2 \tag{2.14}$$

Hence the Pearson residuals are

$$r_{ij}^P = \frac{F_{ij}^P - \hat{f}_j^P}{\hat{\sigma}_j^P} \tag{2.15}$$

and

$$r_{ij}^{Q^{-1}} = \frac{Q_{ij}^{-1} - \hat{q}_j^{-1}}{\tau_j^P}. \tag{2.16}$$

Finally, the estimate of the correlation coefficient is given in equation *(4.15)*.

$$\hat{\rho}^P = \frac{\sum_{i,j} r_{ij}^{Q^{-1}} r_{ij}^P}{\sum_{i,j} \left(r_{ij}^{Q^{-1}}\right)^2}. \tag{2.17}$$

Similarly, for incurred data, the estimates of the development factor parameters can be interpreted as weighted averages of the development factors F_{ij}^I or Q_{ij}, using C_{ij}^I as the weights:

$$\hat{f}_j^I = \frac{\sum_{i=1}^{n-j} C_{i,j+1}^I}{\sum_{i=1}^{n-j} C_{ij}^I} = \sum_{i=1}^{n-j} \frac{C_{ij}^I}{\sum_{i=1}^{n-j} C_{ij}^I} F_{ij}^I \tag{2.18}$$

and

$$\hat{q}_j = \frac{\sum_{i=1}^{n-j} C_{ij}^P}{\sum_{i=1}^{n-j} C_{ij}^I} = \sum_{i=1}^{n-j} \frac{C_{ij}^I}{\sum_{i=1}^{n-j} C_{ij}^I} Q_{ij}. \tag{2.19}$$

Also the unbiased estimates for the variance parameters are defined as follows:

$$\left(\hat{\sigma}_j^I\right)^2 = \frac{1}{n-j-1}\sum_{i=1}^{n-j} C_{ij}^I \left(F_{ij}^I - \hat{f}_j^I\right)^2 . \tag{2.20}$$

and

$$\left(\hat{\tau}_j^I\right)^2 = \frac{1}{n-j}\sum_{i=1}^{n-j+1} C_{ij}^I \left(Q_{ij}^I - \hat{q}_j\right)^2 . \tag{2.21}$$

Hence the Pearson residuals are

$$r_{ij}^I = \frac{F_{ij}^I - \hat{f}_j^I}{\hat{\sigma}_j^I} \tag{2.22}$$

and

$$r_{ij}^Q = \frac{Q_{ij} - \hat{q}_j}{\tau_j^I} . \tag{2.23}$$

And finally, the estimate of the correlation coefficient is given in equation *(2.24)*.

$$\hat{\rho}^I = \frac{\sum_{i,j} r_{ij}^Q \ r_{ij}^I}{\sum_{i,j}\left(r_{ij}^Q\right)^2} . \tag{2.24}$$

Assumptions *A* in section 2.1 have defined the expectations of the development factors, given just the data in the respective triangles. In order to produce a single estimate based on the data from both paid and incurred data, Quarg and Mack (2004) also considers the expectations of the development factors given both triangles and define $E\left[F_{ij}^P \middle| B_{ij}\right] = \lambda_{ij}^P$ and $E\left[F_{ij}^I \middle| B_{ij}\right] = \lambda_{ij}^I$. Using plug-in estimates from equations (4.9) to (4.15), the estimates of the paid MCL development factors are calculated using equation from (4.6):

$$\hat{\lambda}_{ij}^P = \hat{f}_j^P + \hat{\rho}^P \frac{\hat{\sigma}_j^P}{\hat{\tau}_j^P}\left(Q_{ij}^{-1} - \hat{q}_j^{-1}\right). \tag{2.25}$$

Similarly, plug-in estimates from equations (4.16) to (4.22) are used in equation (4.8) so that the estimates of the incurred development factors are

$$\hat{\lambda}_{ij}^I = \hat{f}_j^I + \hat{\rho}^I \frac{\hat{\sigma}_j^I}{\hat{\tau}_j^I}\left(Q_{ij} - \hat{q}_j\right). \tag{2.26}$$

3. Bootstrapping and Claims Reserving

Bootstrapping is a simulation-based approach to statistical inference. It is a method for producing sampling distributions for statistical quantities of interest by generating pseudo samples, which are obtained by randomly drawing, with *replacement*, from observed data. In simple terms, bootstrapping is a re-sampling procedure and all the pseudo samples generated by bootstrapping are subsets of the observed sample or identical to the observed sample.

It should be emphasized that bootstrapping is a method rather than a model. Bootstrapping is useful only when the underlying model is correctly fitted to the data, and bootstrapping is applied to data which are required to be independent and identically distributed. The bootstrapping method was first introduced by Efron (1979) and a good introduction to the algorithm can be found in Efron and Tibshirani (1993).

For the purpose of clarity we begin by giving a general bootstrapping algorithm and briefly reviewing previous applications of bootstrapping to claims reserving. In section 4, we show how an algorithm of this type can be applied to the MCL. Suppose we have a sample $\vec{X}$ and we require the distribution of a statistic $\hat{\theta}$. The following three steps comprise the simplest bootstrapping process:

1 Draw a bootstrap sample $\vec{X}_1^B = \{X_1^B, X_2^B, ..., X_n^B\}_1$ from the observed data $\vec{X} = \{X_1, X_2, ..., X_n\}$.
2 Calculate the statistic interest $\hat{\theta}_1^B$ for the first bootstrap sample $\vec{X}_1^B = \{X_1^B, X_2^B, ..., X_n^B\}_1$.
3 Repeat steps 1 and 2 *N* times.

By repeating steps 1 and 2 *N* times, we obtain a sample of the unknown statistic $\hat{\theta}$, calculated from *N* pseudo samples, i.e. $\vec{\theta}^B = \{\hat{\theta}_1^B, \hat{\theta}_2^B, ..., \hat{\theta}_N^B\}$. When $N \geq 1000$, the empirical distribution constructed from $\vec{\theta}^B = \{\hat{\theta}_1^B, \hat{\theta}_2^B, ..., \hat{\theta}_N^B\}$ can be taken as the approximation to the distribution for the statistic interest $\hat{\theta}$. Hence all the quantities of the statistic interest $\hat{\theta}$ can be obtained since such a distribution contains all the information related to $\hat{\theta}$.

The above bootstrapping algorithm can be applied to the prediction distributions for the best estimates in stochastic claims reserving subject. England and Verrall (2007) contains an excellent review on the application. In addition, Lowe (1994), England and Verrall (1999) and Pinheiro (2003) are also good resources for more details. England and Verrall (2007) showed how bootstrapping can be used for recursive models, following on from the earlier papers (England and Verrall, 1999 and England 2002) which applied bootstrapping to the over-dispersed Poisson model.

It should be noted here that the Pearson residuals are commonly used rather than the original data in the Generalized Linear Model (GLM) framework. The Pearson residuals are required in order to scale the response variables in the GLM so that they are identically distributed. This is necessary because the bootstrap algorithm requires that the response variables are independent and identically distributed.

To our knowledge, there has not been any consideration of bootstrapping for dependent data in the actuarial literature. It should be noted here that a model taking account of all information available could be potentially very valuable, even when the data is dependent in practice. The dependence makes it even difficult to calculate the prediction error theoretically. For these reasons, we believe that adopting bootstrap method for these models is worthy of investigation, particularly in order to obtain the predictive distribution of the estimates of outstanding liabilities.

4. Bootstrapping the Munich chain ladder model

This section considers bootstrapping the MCL model. In section 4.1 we describe the methodology and in section 4.2 we give the algorithm that is used.

4.1 Methodology

The method of bootstrapping stochastic chain ladder models can be seen in a number of different contexts. England and Verrall (2007) categorize the models as recursive and non-recursive and show how bootstrapping methods can be applied in either case. Since we are dealing with recursive models here, we follow England and Verrall and consider the observed development link ratios rather than the claims data themselves. In other words, for Mack's distribution-free chain ladder model the link ratios, F_{ij}, are randomly drawn against C_{ij}, noting that $E\left[F_{ij} \middle| C_{ij}\right] = E\left[\frac{C_{i,j+1}}{C_{ij}} \middle| C_{ij}\right] = f_j C_{ij}$. For this reason, the bootstrap estimates of the development factors f_j^B which are obtained by taking weighted averages of the bootstrapped observed link ratios, F_{ij}^B, use C_{ij} rather than C_{ij}^B as the weights.

However, this method cannot be simply extended to the MCL model, since this model is designed for dealing with two sets of correlated data, the paid and incurred claims. This means that it is not possible to use the normal bootstrap approach: the independence assumption cannot be met any more.

In order to address the problem of how to adapt the existing bootstrap approach in order to cope with the MCL model for dependent data sets, the consideration of the correlation is crucial. It should be noted that the correlation which is observed in the data represents real dependence between the paid and incurred data, and the model is specifically designed because of this dependence. Therefore, it should remain unchanged within any re-sampling procedure. The straightforward solution is to draw samples *pairwise* so that the correlation between the two dependent original data sets will not be broken when generating a sampling distribution for a statistic interest.

Obviously, when bootstrapping the recursive MCL model, the residuals of the paid and incurred link ratios are required instead of the raw data. The question arises of how to deal with these residuals in order to meet the requirement of not breaking the observed dependence between paid and incurred claims,

The answer is to group all the four sets of residuals calculated in the MCL model, i.e. the paid and incurred development link ratios, the ratios of incurred over paid claims from the previous years and its reverse, individually. This is because that the paid claims (incurred claims) are correlated to the incurred claims (paid claims) that are from only the next (previous) year and doing this will preserve the required dependence. And also the correlation coefficient of paid and incurred claims is equal to the correlation coefficient of those residuals, as stated in equations (2.6) and (2.9).

In fact, in the case of the paid claims data, the triangle containing the residuals of the observed paid link ratios and the triangle containing the residuals of the ratios of incurred over paid (except the first column), are paired together, individually, with the same dimensions. And it is the same procedure for the incurred claims data. However, note that the ratios of the paid over incurred claims and the reverse, indicate the same information. Therefore, the ratios should remain unchanged when pairing them with paid and incurred claims with the same dimensions. The consequence of this is that all the four sets of residuals for paid, incurred link ratios and the ratios of incurred over paid claims and the reverse should all be grouped together.

Note here that an alternative approach would be to group three sets of residuals: the residuals of the paid and incurred link ratios and either the residuals of the paid over incurred ratios or the reverse. This would produce the same results as grouping four sets of residuals as the residuals of paid over incurred ratios and the reserve can always be calculated from each other. However, it is simpler to group all the four sets as the calculation of the fourth set of residuals is naturally skipped in this case.

Obviously, this combines the four residuals triangles into one new triangle that consists of these grouped residuals and we name it as the grouped residual triangle. In each unit from this grouped residual triangle, the residuals are from the same accident and development year and correspond to paid and incurred claims. Therefore, the new grouped residual triangle contains all the information available and meanwhile maintains the observed dependence.

When applying bootstrapping, this grouped triangle is considered as the observed sample. And the new generated pseudo samples are obtained by random drawing, with replacement, from this grouped triangle.

The re-sampled incurred and paid triangles can be obtained by separating the pairs in the pseudo sample generated as above and backing out the residual definition. The MCL approach can then be applied to calculate all the statistics of interest for the re-sampled paid and incurred triangles. i.e. the correlation coefficient for paid and incurred, the paid and incurred development factors, the ratios of paid over incurred or the inverse, and the variances. Finally, adjusting the paid and incurred development factors by the correlation coefficient using the MCL approach, the bootstrapped MCL reserve estimates are obtained. This completes a single bootstrap iteration.

Again, the bootstrap method provides only the estimation variance of the MCL model. In order to include the prediction error and estimate the predictive distribution for the MCL estimates of outstanding liabilities, an additional step is added at the end of each of the bootstrap iteration, which is to add the process variance to the estimation variance. In the context of bootstrapping, the way to do this is to simulate new observations from an assumed process distribution with the mean and variance obtained in the same bootstrap iteration. In this paper, an underlying normal distribution is assumed for both the cumulative paid or incurred claims and more details are given in the section 4.2.

In order to obtain a reasonable approximation to the predictive distribution, at least 1000 pseudo samples are required. For each of the pseudo samples, the row totals and overall total of outstanding liabilities are stored so that the sample means, sample variances and the empirical distributions can be calculated and plotted. They are taken as the approximations to the best estimates of outstanding liabilities, the prediction errors and the predictive distributions of the outstanding liabilities. Also, an estimate of any required percentile and confidence interval can be calculated from the predictive distribution.

In order to satisfy the assumption that the sample is identically distributed in the bootstrapping procedure, the Pearson residuals are calculated and used. As in England and Verrall (2007), we use the Pearson residuals of the observed development factors rather than those for the actual claims, since we are using recursive models. Note that a bootstrap bias correction is also needed, and the simplest way to do this is to multiply the residuals by $\sqrt{n/(n-p)}$, where p is the number of parameters estimated in the model and n is the number of residuals.

In addition to drawing the grouped sample for bootstrapping correlated data sets, there are also two other practical points that should be mentioned. The first is to note that the fitted values are obtained by starting from the final diagonal in each triangle and working backwards, by dividing by the development factors. The second is that the zero residuals which appear in both triangles are also left out.

4.2 Algorithm

This section provides the algorithm, step by step, which is needed in order to implement the bootstrap process introduced in section 4.1,

- Apply the MCL model to both the cumulative paid and incurred claims data to obtain the residuals for all the four sets ratios, the paid, incurred link ratios, the paid over incurred ratios and the reverse. They can be obtained from following equations:

$$r_{ij}^{P} = \frac{F_{ij}^{P} - \hat{f}_{j}^{P}}{\hat{\sigma}_{j}^{P}}, \ r_{ij}^{Q^{-1}} = \frac{Q_{ij}^{-1} - \hat{q}_{j}^{-1}}{\tau_{j}^{P}}, \ r_{ij}^{I} = \frac{F_{ij}^{I} - \hat{f}_{j}^{I}}{\hat{\sigma}_{j}^{I}} \text{ and } r_{ij}^{Q} = \frac{Q_{ij} - \hat{q}_{j}}{\tau_{j}^{I}}.$$

- Adjust the Pearson residual estimates by multiplying $\sqrt{n/(n-p)}$ to correct the bootstrap bias, where *p* is the number of parameters estimated in the model and *n* is the number of residuals.

- Group all the four residuals, i.e. r_{ij}^{P}, $r_{ij}^{Q^{-1}}$, r_{ij}^{I} and r_{ij}^{Q} together. We write this as $U_{ij} = \left\{\left(r_{ij}^{P}\right),\left(r_{ij}^{Q^{-1}}\right),\left(r_{ij}^{I}\right),\left(r_{ij}^{Q}\right)\right\}$.

- Start the iterative loop to be repeated N times ($N \geq 1000$). This consists of the following steps:

    1. Randomly sample from the grouped residuals with replacement, denoted as $U_{ij}^{B} = \left\{\left(r_{ij}^{P}\right)^{B},\left(r_{ij}^{Q^{-1}}\right)^{B},\left(r_{ij}^{I}\right)^{B},\left(r_{ij}^{Q}\right)^{B}\right\}$, from the grouped triangle so that a pseudo sample of the grouped residuals is created.

    2. Calculate the MCL reserves for the pseudo samples of paid and incurred claims by fitting the MCL model, as shown below.

        - Calculate the corresponding correlation coefficient for the re-sampled data using the pseudo residuals $\left(r_{ij}^{P}\right)^{B}$, $\left(r_{ij}^{Q^{-1}}\right)^{B}$, $\left(r_{ij}^{I}\right)^{B}$ and $\left(r_{ij}^{Q}\right)^{B}$ as follows,

$$\left(\hat{\rho}^{P}\right)^{B} = \frac{\sum_{i,j}\left(r_{ij}^{Q^{-1}}\right)^{B}\left(r_{ij}^{P}\right)^{B}}{\sum_{i,j}\left(\left(r_{ij}^{Q^{-1}}\right)^{B}\right)^{2}} \text{ and } \left(\hat{\rho}^{I}\right)^{B} = \frac{\sum_{i,j}\left(r_{ij}^{Q}\right)^{B}\left(r_{ij}^{I}\right)^{B}}{\sum_{i,j}\left(\left(r_{ij}^{Q}\right)^{B}\right)^{2}}.$$

        where $\left(r_{ij}^{P}\right)^{B}$, $\left(r_{ij}^{Q^{-1}}\right)^{B}$, $\left(r_{ij}^{I}\right)^{B}$ and $\left(r_{ij}^{Q}\right)^{B}$ denote the re-sampled residuals, which are obtained by simply separating $U_{ij}^{B} = \left\{\left(r_{ij}^{P}\right)^{B},\left(r_{ij}^{Q^{-1}}\right)^{B},\left(r_{ij}^{I}\right)^{B},\left(r_{ij}^{Q}\right)^{B}\right\}$.

        - Calculate the pseudo samples of the four triangles for the paid, incurred link ratios, the ratios of paid over incurred and the reverse by inverting the Pearson residuals definition as follows:

$$\left(F_{ij}^{P}\right)^{B} = \frac{\left(r_{ij}^{P}\right)^{B}\hat{\sigma}_{j}^{P}}{\sqrt{C_{ij}^{P}}} + \hat{f}_{j}^{P},\ \left(Q_{ij}^{-1}\right)^{B} = \frac{\left(r_{ij}^{Q^{-1}}\right)^{B}\hat{\tau}_{j}^{P}}{\sqrt{C_{ij}^{P}}} + q_{j}^{-1},$$

and

$$\left(F_{ij}^{I}\right)^{B} = \frac{\left(r_{ij}^{I}\right)^{B}\hat{\sigma}_{j}^{I}}{\sqrt{C_{ij}^{I}}} + \hat{f}_{j}^{I},\ \left(Q_{ij}\right)^{B} = \frac{\left(r_{ij}^{Q}\right)^{B}\hat{\tau}_{j}^{I}}{\sqrt{C_{ij}^{I}}} + \hat{q}_{j}.$$

- Calculate the $C_{ij}^{P}-$weighted and $C_{ij}^{I}-$weighted average of the bootstrap paid and incurred development factors as follows:

$$\left(\hat{f}_{j}^{P}\right)^{B} = \frac{\sum_{i=1}^{n-j} C_{i,j+1}^{P}}{\sum_{i=1}^{n-j} C_{ij}^{P}} = \sum_{i=1}^{n-j} \frac{C_{ij}^{P}}{\sum_{i=1}^{n-j} C_{ij}^{P}} \left(F_{ij}^{P}\right)^{B},\ \left(\hat{q}_{j}^{-1}\right)^{B} = \frac{\sum_{i=1}^{n-j} C_{ij}^{I}}{\sum_{i=1}^{n-j} C_{ij}^{P}} = \sum_{i=1}^{n-j} \frac{C_{ij}^{P}}{\sum_{i=1}^{n-j} C_{ij}^{P}} \left(Q_{ij}^{-1}\right)^{B}$$

and

$$\left(\hat{f}_{j}^{I}\right)^{B} = \frac{\sum_{i=1}^{n-j} C_{i,j+1}^{I}}{\sum_{i=1}^{n-j} C_{ij}^{I}} = \sum_{i=1}^{n-j} \frac{C_{ij}^{I}}{\sum_{i=1}^{n-j} C_{ij}^{I}} \left(F_{ij}^{I}\right)^{B},\ \left(\hat{q}_{j}\right)^{B} = \frac{\sum_{i=1}^{n-j} C_{ij}^{P}}{\sum_{i=1}^{n-j} C_{ij}^{I}} = \sum_{i=1}^{n-j} \frac{C_{ij}^{I}}{\sum_{i=1}^{n-j} C_{ij}^{I}} \left(Q_{ij}\right)^{B}.$$

Note that the weights used here are from the original data sets and not from the pseudo samples.

- Calculate the bootstrap development factors adjusted by the correlation coefficient between the pseudo samples as follows:

$$\left(\hat{\lambda}_{ij}^{P}\right)^{B} = \left(\hat{f}_{j}^{P}\right)^{B} + \left(\hat{\rho}^{P}\right)^{B} \frac{\hat{\sigma}_{j}^{P}}{\hat{\tau}_{j}^{P}} \left(Q_{ij}^{-1} - \left(\hat{q}_{j}^{-1}\right)^{B}\right)$$

and

$$\left(\hat{\lambda}_{ij}^{I}\right)^{B} = \left(\hat{f}_{j}^{I}\right)^{B} + \left(\hat{\rho}^{I}\right)^{B} \frac{\hat{\sigma}_{j}^{I}}{\hat{\tau}_{j}^{I}} \left(Q_{ij} - \left(\hat{q}_{j}\right)^{B}\right),$$

for the re-sampled bootstrap paid and incurred run-off triangles, respectively.

- Calculate the future cumulative payment using the following equations

$$\left(\hat{C}_{i,n-i+2}^{P}\right)^{B} = \left(\hat{f}_{n-i+1}^{P}\right)^{B} * C_{i,n-i+1}^{P} \text{ and } \left(\hat{C}_{i,j+1}^{P}\right)^{B} = \left(\hat{f}_{j}^{P}\right)^{B} * \left(\hat{C}_{ij}^{P}\right)^{B}$$

and

$$\left(\hat{C}_{i,n-i+2}^{I}\right)^{B} = \left(\hat{f}_{n-i+1}^{I}\right)^{B} * C_{i,n-i+1}^{I} \text{ and } \left(\hat{C}_{i,j+1}^{I}\right)^{B} = \left(\hat{f}_{j}^{I}\right)^{B} * \left(\hat{C}_{ij}^{I}\right)^{B},$$

where $j > n-i+1$.

3. Simulate a future payment for each cell in the lower triangle for both paid and incurred claims, from the process distribution with the mean and variance calculated from previous step. To do this, the following steps are required:

   - For the *one-step-ahead* predictions from the leading diagonal, a normal distribution is assumed. i.e. for $2 \le i \le n$,

$$C^P_{i,n-i+2} \sim Normal\left(\left(\hat{\lambda}^P_{i,n-i+1}\right)^B C^P_{i,n-i+1}, \left(\left(\hat{\sigma}^P_{n-i+1}\right)^B\right)^2 C^P_{i,n-i+1}\right) \text{ for paid claims}$$

   and

$$C^I_{i,n-i+2} \sim Normal\left(\left(\hat{\lambda}^I_{i,n-i+1}\right)^B C^I_{i,n-i+1}, \left(\left(\hat{\sigma}^I_{n-i+1}\right)^B\right)^2 C^I_{i,n-i+1}\right) \text{ for incurred claims.}$$

   - For the *two-step-ahead* predictions up to the *n-step-ahead* predictions, normal distributions are still assumed, but with the mean and variance calculated from previous prediction instead of the observed data. i.e. for $3 \le k \le n$ and $n-k+3 \le j \le n$,

$$C^P_{kl} \sim Normal\left(\left(\hat{\lambda}^P_{k,l-1}\right)^B \hat{C}^P_{k,l-1}, \left(\left(\hat{\sigma}^P_{l-1}\right)^B\right)^2 \hat{C}^P_{k,l-1}\right) \text{ for paid claims,}$$

   and

$$C^I_{kl} \sim Normal\left(\left(\hat{\lambda}^I_{k,l-1}\right)^B \hat{C}^I_{k,l-1}, \left(\left(\hat{\sigma}^I_{l-1}\right)^B\right)^2 \hat{C}^I_{k,l-1}\right) \text{ for incurred claims.}$$

- Sum the simulated payments in the future triangle by origin year and overall to give the origin year and total reserve estimates respectively.

- Store the results, and return to the start of the iterative loop.

5. Examples

This section illustrates the bootstrapping approach to the MCL and uses two numerical examples to assess the results. The first example uses the data from Quarg and Mack (2004). Example 2 uses market data from Lloyd's which have been scaled for confidentiality reasons. These data relate to aggregated paid and incurred claims for two Lloyd's syndicates, categorized at risk level.

Example 1 is included in order to illustrate the results for the original set of data used by Quarg and Mack (2004). The purpose of example 2 is to illustrate that the MCL model does not necessarily provide better results in all situations. The indications from our results that it performs better when the data have less inherent variability and are less "jumpy".

5.1 Example 1

In this section, we use the data from Quarg and Mack (2004). 10,000 bootstrap simulations were carried out, and the data and results are shown in Tables 1 to 5.

Table 1. Paid Claim Data from Quarg and Mack (2004)

576	1,804	1,970	2,024	2,074	2,102	2,131
866	1,948	2,162	2,232	2,284	2,348	
1,412	3,758	4,252	4,416	4,494		
2,286	5,292	5,724	5,850			
1,868	3,778	4,648				
1,442	4,010					
2,044						

Table 2. Incurred Claim Data from Quarg and Mack (2004)

978	2,104	2,134	2,144	2,174	2,182	2,174
1,844	2,552	2,466	2,480	2,508	2,454	
2,904	4,354	4,698	4,600	4,644		
3,502	5,958	6,070	6,142			
2,812	4,882	4,852				
2,642	4,406					
5,022						

Tables 1 and 2 show the data, and in order to observe the run-off nature of the data in a straightforward way, figures 1 and 2 show plots of the data from table 1 and 2, respectively. From figures 1 and 2, it can be seen that the data are stable and not too much spread out.

Figure 1. Paid Claims

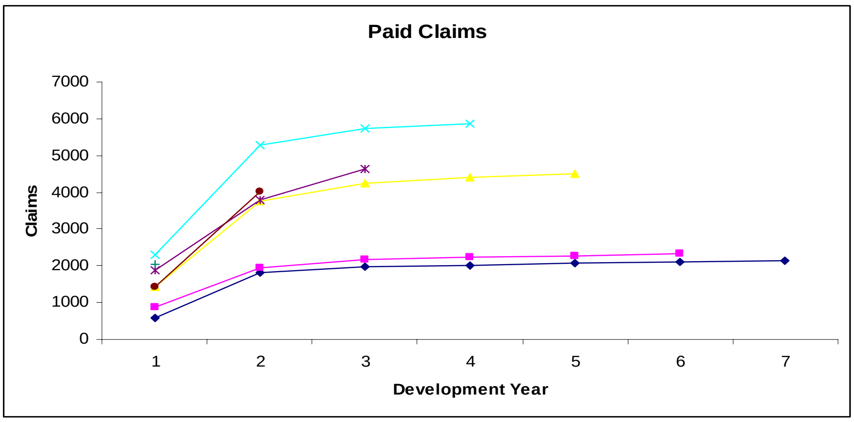
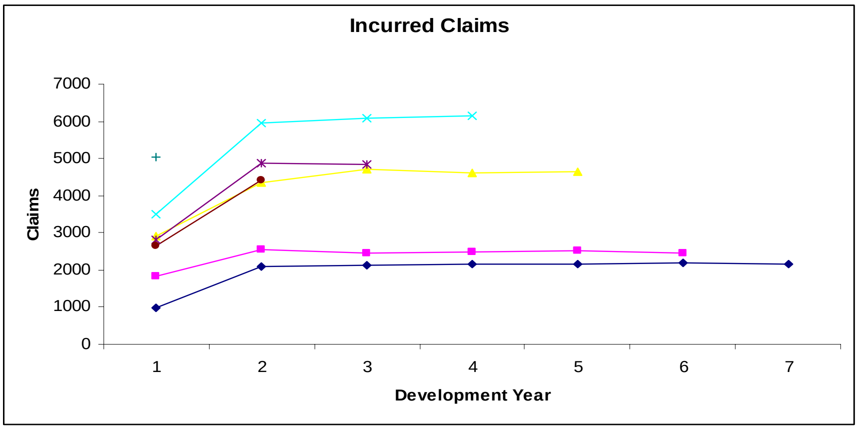


Figure 2. Incurred Claims



The bootstrap methodology described in this paper has been applied, and the results are shown in Tables 3 and 4. Table 3 simply shows that the theoretical MCL reserves (from Quarg and Mack) and the mean of the bootstrap distributions are close to each other in both cases of paid and incurred claims. Table 4 displays the bootstrap prediction error of the MCL reserves projected by both paid and incurred claims. These are displayed both in absolute terms, and as a percentage of the Reserves.

Table 3. Bootstrap Reserves and MCL Reserves

	Bootstrap Reserves		*MCL Reserves*	
	Paid	***Incurred***	***Paid***	***Incurred***
Year 1	0	43	0	43
Year 2	35	95	35	96
Year 3	106	128	103	135
Year 4	275	317	269	326
Year 5	294	287	289	302
Year 6	672	649	646	655
Year 7	5,512	5,655	5,505	5,606
Total	6,893	7,175	6,846	7,163

Table 4. Bootstrap Prediction Error

	Prediction Error		Prediction Error %	
	Paid	Incurred	Paid	Incurred
Year 1	0	0		0%
Year 2	5	5	14%	5%
Year 3	44	67	42%	52%
Year 4	57	84	21%	26%
Year 5	69	99	24%	35%
Year 6	207	204	31%	31%
Year 7	723	695	13%	12%
Total	755	762	11%	11%

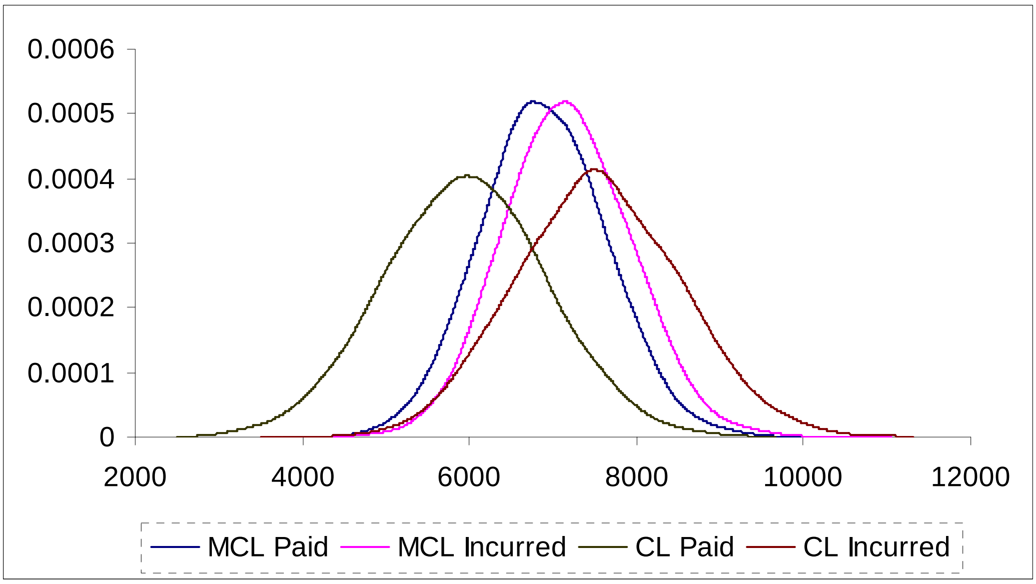
It is interesting to compare the results with those from the chain ladder model. For this, we used the bootstrap results based on the over-dispersed Poisson or over-dispersed negative binomial models described in England and Verrall (2007). Since the purpose of the MCL model is to use more data to improve the estimation of the reserves, it is expected that the prediction errors should be lower than the straightforward CL model. This is confirmed for these data by Table 5, which shows that the prediction error of the MCL reserves is lower than the prediction error of CL reserves. As has been said above, this should not be surprising since the MCL reserves use more information than the CL reserves.

Table 5. Bootstrap Predictions of CL and MCL Models

	Bootstrap CL Prediction Error %		Bootstrap MCL Prediction Error %	
	Paid	Incurred	Paid	Incurred
Year 1	-	0%	-	0%
Year 2	45%	9%	14%	5%
Year 3	33%	96%	42%	52%
Year 4	21%	38%	21%	26%
Year 5	18%	62%	24%	35%
Year 6	31%	47%	31%	31%
Year 7	22%	14%	13%	12%
Total	16%	13%	11%	11%

In figure 3, the distributions of the MCL and CL reserve projections for paid and incurred, are plot respectively in order to compare the results more straightforwardly. Figure 3 shows that the paid and incurred best reserve estimates are very close when using MCL approach. And on the contrast, the paid and incurred best reserve estimates, projected by CL method, are much further compared with the MCL case. More importantly, the CL method provides a much more spread out PDF graph than the MCL approach, both in paid and incurred cases. This means that the MCL is a better method as it not only bridge the gap between paid and incurred best reserves estimates, but also produces a smaller uncertainty around the best reserve estimates.

Figure 3. Predictive Distributions of Overall Reserves
A comparison between CL and MCL projections



5.2 Example 2

In this section, a set of aggregate data from Lloyd's syndicates are considered. In this case, the data are not as stable or well-behaved and the results are quite different. Tables 6 and 7 show the data, which are plotted in figures 4 and 5. It can be seen from these figures that the data are much more unstable and more spread out compared with the previous two examples.

Table 6. Scaled Aggregate Paid Claims at Risk Level from Lloyd's

184	1,845	3,748	5,400	6,231	9,006	9,699	10,008	10,035	10,068
155	1,483	3,768	7,899	8,858	13,795	15,360	15,895	19,333	
676	2,287	10,635	16,102	22,177	28,825	29,828	30,700		
67	367	2,038	2,879	6,329	14,366	16,201			
922	1,693	3,523	4,641	6,431	8,325				
22	488	3,424	5,649	7,813					
76	435	1,980	5,062						
24	1,782	3,881							
39	745								
306									

Table 7. Scaled Aggregate Incurred Claims at Risk Level from Lloyd's

1,530	8,238	10,564	12,332	12,173	10,576	10,630	10,316	10,325	10,280
1,505	6,247	8,728	10,500	15,241	16,720	16,845	16,829	19,675	
2,505	6,150	17,937	22,143	29,511	33,336	32,162	31,500		
204	2,748	9,984	13,167	16,523	17,807	18,959			
2,285	4,361	6,432	8,834	12,092	15,309				
269	5,549	7,214	12,422	13,581					
1,271	2,657	6,187	11,004						
298	3,533	6,423							
2,023	5,415								
1,779									

Figure 4. Scaled Paid Claims from Lloyd's Market

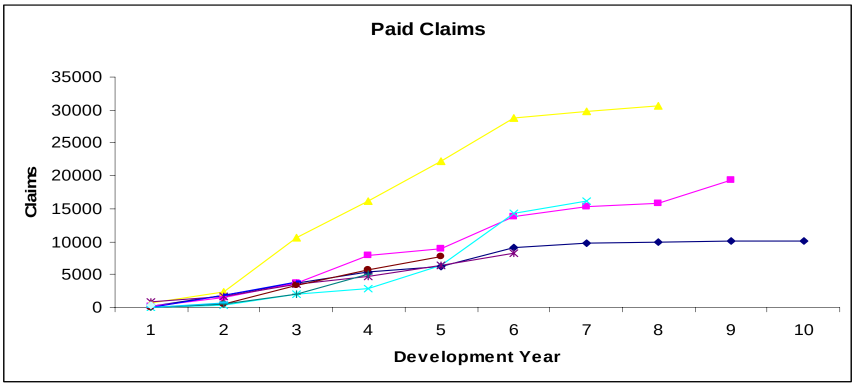
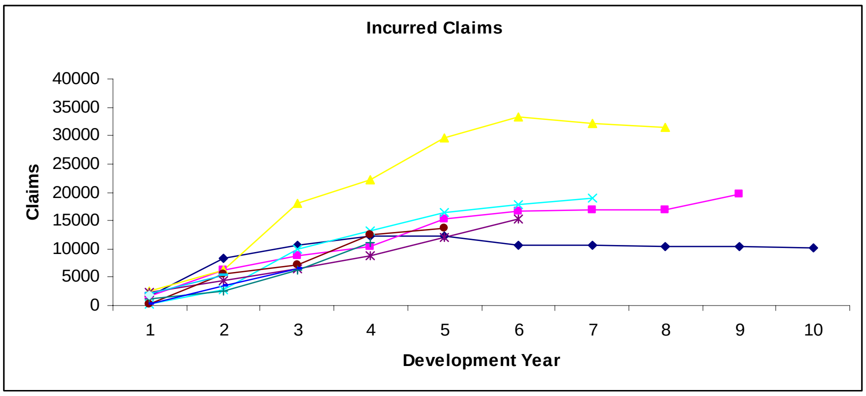


Figure 5. Scaled Incurred Claims from Lloyd's Market



The MCL model still produces consistent ultimate loss predictions for this data set, as shown in table 8. However, the prediction error contained in table 9, estimated by the bootstrap MCL approach, appears to be much higher than for the previous examples.

Table 8. Bootstrap Reserves and MCL Reserves

	Bootstrap Reserves		MCL Reserves	
	Paid	Incurred	Paid	Incurred
Year 1	0	212	0	212
Year 2	48	255	46	258
Year 3	4,177	3,945	3,197	3,974
Year 4	7,319	4,548	6,692	4,306
Year 5	18,366	9,007	17,223	8,200
Year 6	10,708	8,492	10,456	8,314
Year 7	14,291	11,553	14,430	11,219
Year 8	9,670	8,845	9,004	9,051
Year 9	23,980	19,987	23,584	19,185
Year 10	27,901	24,542	28,190	24,633
Total	116,459	91,386	112,822	89,351

Table 9. Bootstrap Prediction Error

	Prediction Error		Prediction Error %	
	Paid	Incurred	Paid	Incurred
Year 1	0	0	-	0%
Year 2	55	234	113%	92%
Year 3	3,489	3,693	84%	94%
Year 4	5,630	3,059	77%	67%
Year 5	18,102	6,905	99%	77%
Year 6	16,346	5,831	153%	69%
Year 7	18,711	7,320	131%	63%
Year 8	17,507	9,291	181%	105%
Year 9	31,947	14,192	133%	71%
Year 10	51,698	36,989	185%	151%
Total	86,509	47,804	74%	52%

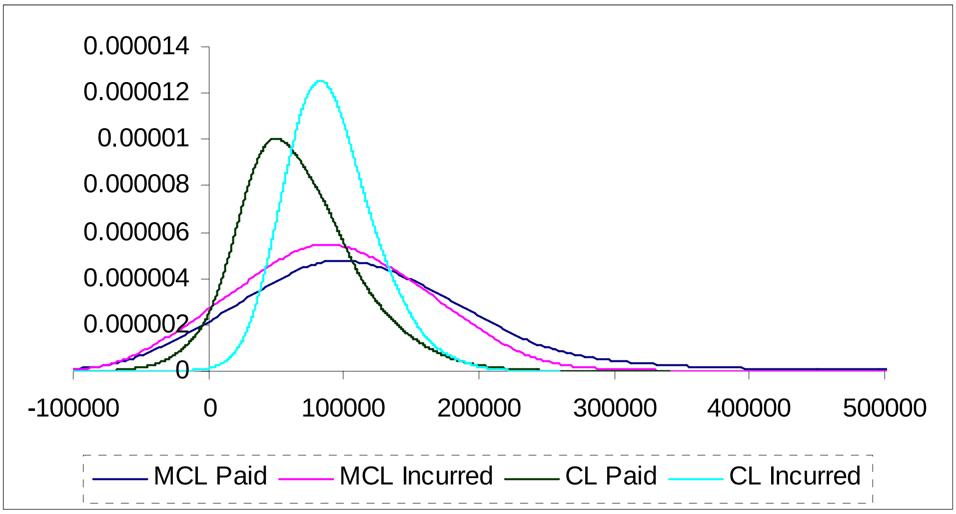
This becomes even clearer when comparing with the chain ladder technique. Table 10 shows that the MCL model produces a higher prediction error than the CL model. The conclusion from this is that although the MCL model uses more data, and should be expected to produce lower prediction errors, this is not always the case in practice. We believe that the reason for this is that the assumptions made by the MCL model – the specific dependencies assumed – are not as strong as expected in this case. A conclusion from this is that the data have to be examined carefully before the MCL model is applied.

Table 10. Bootstrap Predictions of CL and MCL Models

	Bootstrap CL Prediction Error %		Bootstrap MCL Prediction Error %	
	Paid	Incurred	Paid	Incurred
Year 1	-	0%	-	0%
Year 2	148%	164%	113%	92%
Year 3	102%	88%	84%	94%
Year 4	99%	57%	77%	67%
Year 5	92%	28%	99%	77%
Year 6	71%	37%	153%	69%
Year 7	72%	36%	131%	63%
Year 8	62%	49%	181%	105%
Year 9	97%	55%	133%	71%
Year 10	185%	103%	185%	151%
Total	66%	36%	74%	52%

Again, in figure 6, a more straightforward comparison is provided. It can be seen that the CL model produces less variable forecasts than the MCL model for this set of data: the uncertainty or the prediction variance are relatively smaller than the MCL approach.

Figure 6. Predictive Distributions of Overall Reserves
A comparison between CL and MCL projections



6. Conclusion

This paper has shown how a bootstrapping approach can be used to estimate the predictive distribution of outstanding claims for the MCL model. The model deals with two dependent data sets, i.e. the paid and incurred claims triangles, for general insurance reserving purposes. We believe that bootstrapping is well-suited for these purposes from a practical point of view, since it avoids complicated theoretical calculations and is easily implemented in a simple spreadsheet. This paper adapts the method by taking account of the dependence observed in the data and maintaining it by re-sampling pairwise.

A number of examples have been given, which show that the MCL model does not always produce superior results to the straightforward chain ladder model. As a consequence, we believe that it is important for the data to be carefully checked to test whether the dependency assumptions of the MCL model are valid for each data set before it is applied.

References

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

Efron, B. and Tibshirani R.J. (1993). An introduction to the bootstrap. *New York: Chapman and Hall*.

England, P.D. and Verrall, R.J. (1999). Analytic and bootstrap estimates of prediction errors in claims reserving. *Insurance: Mathematics and Economics* **25**, 281-293.

England, P.D. and Verrall, R.J. (2002) Stochastic Claims Reserving in General Insurance (with discussion). *British Actuarial Journal* **8**, pp 443-544

England, P.D. (2002). Addendum to "Analytic and bootstrap estimates of prediction errors in claims reserving". *Insurance: Mathematics and Economics* **31**, 461-466.

England P.D. and Verrall, R.J. (2007). Predictive Distributions of Outstanding Liabilities in General Insurance. *Annals of Actuarial Science Institute of Actuaries.* **1**(2), 221-270.

Lowe, J. (1994). A practical guide to measuring reserve variability using: Bootstrapping, Operational Time and a distribution free approach. *Proceedings of the 1994 General Insurance Convention, Institute of Actuaries and Faculty of Actuaries.*

Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin* **23** (2), 214-225.

Murphy, D.M. (1994). Unbiased Loss Development Factors. *Proc. CAS* **81**, 154-222.

Pinheiro, PJR, Andrade e Silva, JM, Centeno, MLC (2003). Bootstrap methodology in claim reserving. *Journal of Risk and Insurance*, **70** (4), 701-714.

Quarg, G. and Mack, T. (2004). Munich Chain Ladder. *Blatter DGVFM* **26**, 597–630.