Continuous Chain Ladder

María Dolores Martínez Miranda
University of Granada, Spain
mmiranda@ugr.es
Jens Perch Nielsen
Cass Business School, City University, London, U.K.
Jens.Nielsen.1@city.ac.uk, festinalente@nielsen.mail.dk
Stefan Sperlich
Universitè de Genève, Département des sciences économiques
Bd du Pont d'Arve 40, CH - 1211 Genéve 4
Stefan.Sperlich@unige.ch
Richard Verrall
Cass Business School, City University, London, U.K.
R.J.Verrall@city.ac.uk

February 11, 2013

Abstract

To be written

Keywords: Chain Ladder; Claims Reserves; Reserve Risk; Multiplicative bias correction; Density estimation, Crossvalidation; smoothing; kernel.

1 Introduction

A crucial element of non-life insurance is to set adequate reserves aside for liabilities that are not fully known, and this is of great importance when considering the risk of insolvency and the capital requirements for non-life insurers. Determining the expected profit or loss in a non-life insurance business is of growing importance because of the Solvency II regulations. For an insurer operating in the non-life insurance business, the ultimate claims amount of an accident year is often not known at the end of that year. It will depend on the business line: for instance, in liability insurance it may be expected that the claims settlement will last several years because of bodily injuries and/or long legal processes. Also the possible time lag between the occurrence of the accident and the manifestation of the consequences of the event may cause a delay. The aim of claims reserving is to estimate the outstanding claims and thereby enable the company to set its reserves. To forecast these outstanding claims, a simple but generally accepted algorithm is the classical chain ladder method (CLM).

A key feature of the vast majority of claims reserving methids used in practice, including the CLM, is that they assume that the data have been aggregated. This aggregation is often done by year, but it could also be done by quarter or month. Whatever time period is used, the key point to note is that this aggregation implies some pre-smoothing of the data. Claims reserving methods which use continuous models (parametric or non-parametric) have been suggested in the literature but they almost invariably involved the use of aggregated data which has therefore been reduced to discrete time data. The approach of this paper is different in that it does not assume that the data have been aggregated: it uses data recorded in continuous time. As this is a new approach, we present methods which are close to the CLM, but we keep them as straightforward as possible. It would be possible to add sophistication to these, but this would detract from the simplicity of the presentation and we leave this for future work. Because the appraoch uses data recorded in continuous time and is based on the philosophy underlying the CLM, it can be viewed as a continuous version of the CLM. For this reason, we have named it " Continuous Chain Ladder".

Methods based on data recorded over a continuous time scale have previously been suggested in the actuarial literature (e.g. Norberg, 1986), but the papers only rarely addressed the implementation of these methods. In the practical context the results have generally been somewhat disappointing. We believe that this somewhat disappointing outcome for continuous reserving so far is that the methods have been too distant from well-known methods such as the CLM to appeal to actuaries. This paper therefore aims to show how reserving using data recorded in continuous time can be viewed as a natural transition from CLM to sophisticated modern statistical methods.

Using the technical language of non-parametric smmothing, this paper will show that the classical CLM can be regarded as a structured histogram on a triangle (Jones 1989). The original CLM groups the data and develops a multiplicative histogram model: we will simply follow the same approach without grouping the data. When the data are not grouped, it can be treated as having a density on the triangle. This multivariate density can be estimated using local linear smoothing methods, and it can then be approximated on the triangle by a multiplicative density. It is also possible to relate this approach to regression. Linton and Nielsen (1998) showed that their method of marginal regression (Linton and Nielsen, 1995), simply minimized the distance from a multivariate regression function by two additive components. Mammen, Linton and Nielsen (1999) developed smooth additive regression as a kernel weighted projection of the data on an additive subspace on non-parametrically defined functions. This research led to a number of extensions including the generalised additive nonparametric smooth backfitting of Yu, Park and Mammen (2008). This is related to the approach taken in this paper, because the multiplicative model can be regarded as a generalized additive model.

The paper is structured as follows...

2 The classical model for aggregated data

One of the most popular method used in reserving is the classical chain ladder method (CLM) which uses simple aggregated triangular arrays of data. Several stochastic models for CLM have been formulated which model directly the aggregated data (see for example Verrall and England 2002, or Wüthrich and Merz, 2008). This paper considers a new and modern approach based on smoothing methods which will provide a different perspective on claims reserving, which gives more insights into classical reserving methods such as the CLM. This section describes briefly the classical formulation which will be a benchmark throughout the paper.

In classical reserving methods it is assumed that the available information is a run-off triangle with dimension m, i.e. a triangle with m rows. Thus, the information is provided in an aggregated way where, in theory, any aggregation periods could be considered, such as quarters, years etc. Depending on the data being considered, each cell in the triangle could contain the number of reported or paid claims (counts data) or aggregated payments (reported or paid). Traditional methods such as the CLM are often applied to all of these types of triangles, with different distributional assumptions used as appropriate. For example, a Poisson model would be suitable for counts and an over-dispersed Poisson for aggregated payments. This paper considers only counts in order to make the density approach as clear as possible. The extension to payments data will be considered in future work.

We assume, without loss of generality, that the data are available as a triangle, and denote the set indexing the periods for which the data are available by $\mathcal{I}_m = \{(i,j) : i = 1,\dots,m, j = 1,\dots,m;\ i+j-1 \leq m\}$. Here i denotes the origin period and j the delay period (i.e. $j-1$ periods delay from i). The aggregated incurred counts triangle can then be written by $\aleph_m = \{N_{ij} : (i,j) \in \mathcal{I}_m\}$, where N_{ij} is the total number of claims of insurance incurred in period i, which are reported in period $i+j-1$ i.e. with $j-1$ periods delay from year i. An example of such aggregated data is shown in the top graph in Figure 1. This data set corresponds to numbers of reported claims (incurred counts) observed for claims that were incurred in the $m = 19$ past years.

It is usual that data in this triangular form are used for classical reserving methods and it is possible to predict the outstanding claims and construct a reserve starting from aggregated data. In order to do this, the methods produce projections in the lower and unexperienced triangle $\mathcal{J}_m = \{(i,j) : i = 2, \ldots, m, j = 1, \ldots, m;\ i + j - 1 > m\}$. The traditional CLM projections can be derived from maximum likelihood estimation under a Poisson model for the aggregated data (see Kuang, Nielsen and Nielsen, 2009). Note that such a model is often quite a reasonable model for counts data. Thus, it is assumed that the cells in the triangle are independently Poisson distributed with cross-classified mean, which is specified by the following multiplicative parametrization:

$$\mathrm{E}[N_{ij}] = \alpha_i \beta_j, \quad (i,j) \in \mathcal{I}_m. \tag{1}$$

By solving the well-known identification problem for this kind of method (see Kuang, Nielsen and Nielsen, 2009), standard tools from generalized linear models provide estimates for the parameters α_i and β_j, for $i, j = 1, \ldots, m$. From these estimates the predicted outstanding claims for each underwriting period are obtained by summing the predicted values for the claims in the lower triangle by row. Also, outstanding claims for future calendar period can be predicted by summing the diagonals in the lower triangle. Both calculations are the common output required from reserving methods. An illustration of the classical chain ladder method on real insurance data is given in Section 5. The next section introduces the continuous approach which underlies the new method proposed in this paper.

2.1 A regression view of the density estimation problem

As indicated above, the aim of this paper is firstly to reformulate the reserving problem in terms of a multivariate density estimation problem and then to develop kernel methods to estimate nonparametrically this density. The classical CLM provides one approach to this non-parametric density estimation problem as is explained in this subsection. This new perspective on further understanding classical reserving methods such as the CLM can provide greater understanding and may also be the key to makel progress in developing modern and powerful methods.

Several papers in the statistical literature have described the connection between the density and the regression problems. This connection has motivated new density estimation methods such as local linear estimator, which will be applied in this paper to the reserving problem. We begin by describing the connection as Fan and Gijbels (1996, pp. 50) did in the unidimensional setting, and consider first the simple univariate scenario to make the exposition more straightforward. Let $X_1, \ldots, X_n$ be a random i.i.d. sample from a population X with continuous density f. Let $a_1 < \cdots < a_m$ denote equally spaced grid points defining $m-1$ contiguous intervals or bins $B_j = (a_j, a_{j+1}]$, for $j = 1, \ldots, m-1$. The extreme points a_1 and a_m are typically chosen in such a way that the support of f is contained in $(a_1, a_m]$. Let Λ_m denote the bin size ($\Lambda_m = a_2 - a_1$). Also for each $j = 1, \ldots, m$, let x_j denote the bin center ($x_j = (a_{j+1} + a_j)/2$), and N_j the bin count defined as the number of data falling in the interval B_j. It is clear that N_j follows a binomial distribution with size parameter n and success probability $p_j = \int_{B_j} f(x)dx$. Therefore when $m \to \infty$ or equivalently $\Lambda_m \to 0$ we have the following approximations:

$$\mathrm{E}\left[\frac{N_j}{n\Lambda_m}\right] \approx f(x_j)$$

and

$$\mathrm{V}\left(\frac{N_j}{n\Lambda_m}\right) \approx \frac{f(x_j)}{n\Lambda_m},$$

for $j = 1, \ldots, m$. Thus the density problem is equivalent to a heteroscedastic regression problem based on the data, $\{(x_j, N_j/\Lambda_m n), j = 1, \ldots, m\}$, which are approximately independent. Equivalently the regression model can be written for the bin counts N_j as

$$N_j = r(x_j) + \varepsilon_j, \tag{2}$$

with $r(\cdot) = n\Lambda_m f(\cdot)$ being the regression function. From a regression estimate $\widehat{r}(\cdot)$ the target density can be estimated as $\widehat{f}(\cdot) = r(\cdot)n\Lambda_m$.

Now if we move to the two-dimensional scenario it can clearly be seen that the classical chain ladder method approaches the density problem through the regression formulation in equation (2). Specifically, the CLM estimates a two-dimensional density f supported in the triangle $\mathcal{I}_m$ and using a multiplicative structure. The bins are constructed as squares of the form $B_{ij} = (a_i, a_{i+1}] \times (b_j, b_{j+1}]$, for $i, j \in \mathcal{I}_m$, with bin length $\Lambda_m = a_2 - a_1 = b_2 - b_1$ being constant. Thus, from the bin counts N_{ij} (the number of data falling in B_{ij}) the regression problem can be formulated as

$$N_{ij} = r(z_{ij}) + \varepsilon_{ij}, \tag{3}$$

based on the data $\{(z_{ij}, N_{ij}, i, j \in \mathcal{I}_m\}$. The two-dimensional covariate $z_{ij} = (x_i, y_j)$, is defined such that x_i and x_j are the midpoints of the intervals $(a_{1,i}, a_{1,i+1}]$ and $(a_{2,j}, a_{2,j+1}]$, respectively, for $i, j \in \mathcal{I}_m$. The regression function is then related with the density by $r(\cdot, \cdot) = n\Lambda_m f(\cdot, \cdot)$. By assuming that the unknown regression function r is multiplicative i.e. $r(\cdot, \cdot) = r_1(\cdot)r_2(\cdot)$, the problem can be solved using classical generalized linear models (GLM) with the logarithm as the link function and some specified error distribution (see for example England and Verrall, 2002, for a description of this approach). The classical Poisson (for claim counts) as defined by equation (1) relies on the Poisson approximation of the binomial distribution. Clearly each bin count, N_{ij}, follow a binomial distribution with parameters n and p_{ij}, where n is the total ultimate number of claims for each accident year and $p_{ij} = \int_{B_{ij}} f(z)dz$. It is well known that the binomial distribution can be approximated by a Poisson distribution: $N_{ij} \hookrightarrow P(np_{ij})$, which justifies a GLM model with log link function and Poisson error distribution. Note that the larger n and the smaller p_{ij} the better is the approximation and therefore the expected performance of the classical CLM.

The above description shows how CLM focuses on the regression approach when considering the estimation of a density and thus it can only work on binned data. The following section describes methods which aim to estimate the underlying density and which are therefore more suited to the consideration of individual data using continuous time.

3 The continuous density approach

Once the reserving problem has been formulated in terms of a bivariate density estimation problem, several powerful methods from modern statistics can assist in providing good solutions. Note that the approach now changes from the regression perspective which is useful when the data are given in an aggregated way (see Appendix A), to a continuous approach where the target function is a continuous two-dimensional density function. As was discussed in the previous section the chain ladder method is defined from the regression perspective, which also starts from a histogram or a binned version of the data. This is also the case with many other approaches in reserving that attempt to introduce smoothing ideas. Some of these approaches are examined in Section 4. This section focuses on the aim of this paper, which is a continuous version of the CLM. This provides solutions for the reserving problem with better statistical properties and which can be extended to the context of individual claims data. Subsection 3.1 describes how a simple kernel density estimator can be developed from the naive and inefficient histogram, which is the basis of classical CLM. A sequence of steps to improve the local linear and multiplicative bias correction estimators can then be introduced, which lead to the proposal for a continuous CLM.

3.1 From the histogram to kernel smoothing

Kernel methods for density estimation arise in an intuitive and natural way from the naive histogram estimator. The application of these kernel methods in reserving relies on the recognition that the classical CLM consists of the construction of a structured histogram on a triangle. A histogram is the simplest nonparametric approach to estimate a density function. The histogram separates the data into distinct non-overlapping *bins*, and constructs bars (hypercubes) with heights defined as the proportion (or the number) of observations falling into each bin. This proportion gives an estimate of the probability density function at the mid point of the bin (see subsection 2.1). As in Section 2.1 we start from the simpler univariate scenario and extend this afterwards to the two-dimensional situation. Consider again a random sample, $X_1, \ldots, X_n$, from a population X with a continuous density f. Consider $m-1$ contiguous intervals $B_j = (a_j, a_{j+1}]$ or bins with bin length Λ_m, which divide the support of f, and let x_j be the midpoints, for $j = 1, \ldots, m-1$. The height of the bar of the histogram with base B_j provides an estimate of the probability density function at the midpoint, x_j. Thus, an estimator of the density f at any point x_0 in the support of f can be derived from the limit concept of ratio between probability mass in a neighborhood of a point and the size of the neighborhood. Using proper mathematics it is an application of the mean value theorem of integral calculus, which implies that

$$\lim_{\Lambda_m \to 0} \frac{P(X \in B_j)}{\Lambda_k} = f(x) \quad \text{if} \;\; x \in B_j \; (j = 1, \ldots, m-1).$$

From this expression the histogram estimator at any point x_0 in the support is defined by

$$\widehat{f}_{hist}(x_0) = \frac{n^{-1} \sum_{i=1}^{n} I\{X_i \in B_j\}}{\Lambda_m} \quad \text{if} \;\; x_0 \in B_j \; (j = 1, \ldots, m).$$

Note that the histogram is not a continuous function, but has jumps at the grid points and has zero derivative everywhere else. This gives estimates which are not only aesthetically undesirable, but, more seriously, could provide to an untrained observer with a misleading impression. In fact, the shape of the histogram can potentially be influenced by where the bin centres are placed. From the above formulation, these are defined by the choice of the width Λ_m and the location of the first point a_1. Partly to overcome this difficulty, and partly for other technical reasons (see for example Silverman, 1986, for a further explanation), it is of interest to consider more sophisticated estimators which can overcome these drawbacks, such as kernel methods.

Moving now to the two-dimensional scenario, assume that $Z_1, \ldots, Z_n$ is an i.i.d. random sample from a population $Z = (X, Y)$, having bivariate continuous density f. Consider a split of the support of f into squares of the form $B_{ij} = (a_{1,i}, a_{1,i+1}] \times (a_{1,j}, a_{2,j+1}]$, with constant length of the sides $\Lambda_m = a_{1,2} - a_{1,1} = a_{2,2} - a_{2,1}$. Following analogous arguments to those used in the univariate case, the simple histogram estimator can be defined at any point $z_0 = (x_0, y_0)$ in the support of f by

$$\widehat{f}_{hist}(z_0) = \nu(z_0)/n\Lambda_m^2,$$

where $\nu(z_0)$ is the number of sample data falling in the square which contains z_0. From this formulation the typical kernel density estimator can be seen as a moving histogram which defines the bins centered at each point where the density is estimated. In this case the bins can overlap and the data falling in the bin are given different weights according to their proximity to the estimation point. Thus, the kernel estimator overcomes the problem of the naive estimator concerning the location of the bins but also it provides a smooth estimate for the target continuous density. The simplest kernel density estimator is the multivariate extension of the Parzen-Rosenblatt estimator. For any estimation point the support, $z_0 = (x_0, y_0)$ this is defined as

$$\widehat{f}_h(z_0) = |h|^{-1} \sum_{i=1}^{n} \mathcal{K}_h(z_0 - Z_i), \tag{4}$$

where $\mathcal{K}_h(\cdot)$ is a two-dimensional kernel and $h = (h_1, h_2)^t \in I\!R_+^2$ is a bandwidth parameter with $|h| = h_1 h_2$. Here we use a simple multiplicative kernel given by $\mathcal{K}_h(u, v) = K_{h_1}(u) K_{h_2}(v)$ with $K_{h_1}(u) = h_1^{-1} K(u/h_1)$ and K being a unidimensional symmetric probability density function. This multiplicative structure is the more convenient for our purposes in the paper, but other general kernels such as spheric kernels could also be considered, and also more general bandwidth parameter such as matrices (see for example Wand and Jones, 1995, pp.103).

Simple kernel methods suffer from well-known boundary problems and further corrections have been suggested in the literature to overcome these problems. Reserving can be viewed as a typical density estimation problem in the boundary region defined by the claims triangle. The triangular support requires refined boundary corrections methods to be considered, and local linear density estimation can be useful in this context. Subsection 3.2 starts by considering the local linear approach to provide an estimator of the density in the triangle. Since the aim is to predict the density in the whole square which includes the future in the lower triangle, we will next assume a multiplicative structure and use the marginal integration method by Linton and Nielsen (1995) to provide a multiplicative local linear estimator for the density. This provides the required predictions in the forecast set (the lower triangle in the square).

3.2 The unstructured local linear estimator in the observation triangle

Nielsen (1999) extended the principle of local linear estimation by Lejeune and Sarda (1992) and Jones (1993) to nonparametric multivariate density estimation with arbitrary boundary regions. Let f denote a two-dimensional density having support in the triangle $\mathcal{I} = \{z = (x, y)^t | 0 \leq x, y \leq T, x + y \leq T\}$ with any $T > 0$. Here for simplicity we assume the origin period is equal to zero. Nielsen's local linear estimator is defined at each point $z_0 = (x_0, y_0)^t \in \mathcal{I}$ as the solution $\widehat{\Theta}_0$ of the following minimization problem:

$$\begin{pmatrix} \widehat{\Theta}_0 \\ \widehat{\Theta}_1 \end{pmatrix} = \arg\min \left\{ \lim_{b \to 0} \int_{\mathcal{I}} \left[\widetilde{f}_b(z) - \widehat{\Theta}_0 - \widehat{\Theta}_1^t (z_0 - z) \right]^2 \mathcal{K}_h(z - z_0) dz \right\}, \quad (5)$$

where $\widetilde{f}_b(z) = n^{-1}(b_1 b_2)^{-1} \sum_{i=1}^n \mathcal{K}_b(z - z_0)$ is the standard kernel estimator in (4) at the point z with bandwidth parameter $b = (b_1, b_2)^t \in I\!R_+^2$ and two-dimensional kernel $\mathcal{K}$. As above in section 2.1 a multiplicative kernel, $\mathcal{K}(x, y) = K(x)K(y)$ is used, with K being a unidimensional symmetric kernel function.

Note that this estimator is only defined in the observation triangle and therefore is not suitable for forecasting purposes. Recall that the forecast horizon is giving by $\mathcal{J} = \{z = (x, y)^t | 0 \leq x, y \leq T, x + y > T\}$. In the next section we assume a multiplicative structure for the density and provide an estimator which can be used to provide forecasts in the lower triangle $\mathcal{J}$.

3.3 The structured local linear density estimator

Now we assume that the target density in the whole square $\mathcal{S} = \{z = (x, y)^t | 0 \leq x, y \leq T\}$ is multiplicative i.e. $f(x, y) = f_1(x) f_2(y)$. The marginal integration method introduced by Linton and Nielsen (1995) can be extended to the density estimation problem in 3.2 through the following two-step method:

> Step 1. From the available data estimate the two-dimensional density in the observation set $\mathcal{I}$ by an estimator $\widehat{f}_h^{\mathcal{I}}(x, y)$, such as the local linear estimators resulting from (5).

Step 2. Assume that the target density is multiplicative, $f(x,y) = f_1(x)f_2(y)$ and estimate f_1 and f_2 through the following minimisation:

$$\min_{f_1,f_2} \int_{\mathcal{I}} \left(\widehat{f}_h^{\mathcal{I}}(x,y) - f_1(x)f_2(y) \right)^2 dxdy. \qquad (6)$$

The minimization at Step 2 can be performed using an iterative algorithm such as the following:

1. Consider an initial estimator of the component f_1 denoted by $\widehat{f}_1^{(0)}$. Let $\widehat{f}^{(0)}$ denote the unstructured estimator for the density in $\mathcal{I}$ defined in Step 1 above.
2. Using $\widehat{f}_1^{(0)}$, $f(x,y) \approx \widehat{f}_1^{(0)}(x)f_2(y)$ so that
$$\int_{\mathcal{I}_y} f(x,y)dx \approx f_2(y) \int_{\mathcal{I}_y} \widehat{f}_1^{(0)}(x)dx$$
with $\mathcal{I}_y = \{x|(x,y) \in \mathcal{I}\}$. Estimate the density f_2 by
$$\widehat{f}_2^{(1)}(y) = \frac{\int_{\mathcal{I}_y} \widehat{f}^{(0)}(x,y)dx}{\int_{\mathcal{I}_y} \widehat{f}_1^{(0)}(x)dx}.$$
3. Using $\widehat{f}_2^{(1)}$, calculate the updated estimator for f_1 by
$$\widehat{f}_1^{(1)}(x) = \frac{\int_{\mathcal{I}_x} \widehat{f}^{(0)}(x,y)dy}{\int_{\mathcal{I}_x} \widehat{f}_2^{(1)}(y)dy}.$$
4. Repeat steps 2-3 until the desired convergence criterion is achieved.

This provides estimates for any point in the square $\mathcal{S} = \{z = (x,y)^t | 0 \leq x,y \leq T\}$. The only requirement in practice is to choose the kernel function K and a bandwidth parameter h which introduces a suitable level of smoothing. The first choice is usually made for practical or theoretical reasons and it usually has a minor impact on the performance of the estimator (see for example Wand and Jones, 1995, for more details about the choice of kernel). However, the choice of the bandwidth parameter can significantly affect the performance of the kernel estimator and because of its importance, it is considered in subsection 3.5.

3.4 A multiplicative bias correction

In this section, we consider a second improved kernel estimator using bias reduction techniques. It is well-known that kernel methods such as those proposed above provide biased estimates. In the context of this paper, bias should be corrected since it could lead to incorrect reserves with serious consequences for the solvency of non-life insurers. Note that the variability is not so relevant because the insurer is usually interested in aggregated values of the estimates such as the total reserves for the future calendar years or even the overall total in the range of years under consideration.

There are several alternative bias reduction methods to correct the kernel estimates at points of large curvature. Here we consider the multiplicative bias correction (MBC) method proposed by Jones, Linton and Nielsen (1995) for univariate density estimation. The estimator is again introduced in two steps. Firstly a multiplicative bias corrected estimator for the density in the observation triangle is defined, which is an unstructured MBC estimator. Secondly, the marginal integration method is applied to provide the structured MBC density estimator.

Consider the unstructured local linear estimator defined in subsection 3.2. Denote this estimator by $\widehat{f}^{\mathcal{I}}_{LL,h}$ and recall that it is supported in $\mathcal{I}$. The unstructured MBC estimator is defined (having the same support) from the following expression:

$$\widehat{f}^{\mathcal{I}}_{MBC,h}(z) = \widehat{f}^{\mathcal{I}}_{LL,h}(z)\widehat{g}^{\mathcal{I}}_{LL,h}(z) \tag{7}$$

where $\widehat{g}^{\mathcal{I}}_{LL,h}$ is the local linear estimator of the ratio $f(z)/\widehat{f}^{\mathcal{I}}_{LL,h}(z)$ obtained by minimising the expression below in Ψ_0.

$$\min_{\Psi_0,\Psi_1}\left\{\lim_{w\to 0}\int_{\mathcal{I}}\left[\widetilde{g}(z;w)-\Psi_0-\Psi_1^t(z_0-z)\right]^2\left(\widehat{f}^{\mathcal{I}}_{LL,h}(z)\right)^{-1}\mathcal{K}_h(z-z_0)dz\right\} \tag{8}$$

Now from the estimator defined in (7) and using a similar method to that described in subsection 3.3 the MBC estimators of the univariate densities f_1 and f_2 are obtained, together with the structured MBC estimator as their product.

3.5 Choice of degree of smoothing

To make the kernel methods presented above aoolicatble in practice, it is necessary to make a suitable choice of the bandwidth or smoothing parameter, and for this optimality criterion is often used. However, for any kernel estimate the theoretically optimal bandwidths are unfeasible in practice and therefore it is necessary to provide a reasonable data-driven bandwidth estimate. The problem of bandwidth selection also arises also when other estimators such as smoothing splines are used. In this case the smoothing parameter defines the appropriate weightings of fit andd smmothness. Section 4 describes some of the approaches that have been previously considered in the reserving literature.

For the kernel density estimators defined above, the bandwidth is a two-dimensional parameter $h = (h_1, h_2)$ which controls the degree of smoothing in each direction. Specifically h_1 and h_2 move between 0 and infinity, thereby corresponding to extreme cases of undersmoothing and oversmoothing, respectively. h_1 defines the degree of smoothing in the underwriting direction and h_2 in the development direction.

There are several possible methods suggested in the literature to choose the bandwidth for a two-dimensional density. One of the simplest and most commonly used is the crossvalidation method (see for example Wand and Jones 1995). The crossvalidation method is an in-sample technique which aims to estimate the optimal bandwidth for the estimator using the sample data. In this paper, two candidates are suggested to estimate the density from a sample in the reserving triangle using a multiplicative structure. These are the local linear estimator and a multiplicative bias-corrected version. We propose here simply to use crossvalidation to find good unstructured density estimators in the observed triangle, and from such estimators to calculate the corresponding structured densities following the method described in Section 3.3.

For either of the unstructured density estimators defined in the triangle $\mathcal{I}$, the LL estimator ($\widehat{f}^{\mathcal{I}}_{LL,h}$) or the MBC estimator ($\widehat{f}^{\mathcal{I}}_{MBC,h}$), the cross-validation score is defined by

$$\mathrm{CV}(h) = \int_{\mathcal{I}} \widehat{f}^{\mathcal{I}}_h(z)^2 dz - 2\sum_{i=1}^{n} \int_{\mathcal{I}} \widehat{f}^{\mathcal{I},[-i]}_h(z) d\widetilde{F}_n(z), \tag{9}$$

with $\widehat{f}^{\mathcal{I},[-i]}_h$ being the leave-one-out version of the estimator $\widehat{f}^{\mathcal{I}}_h$, and $\widetilde{F}_n$ being the empirical distribution function from the sample. The cross-validation bandwidth is then defined as the minimizer of the above CV score. The crossvalidation multiplicative bias-corrected estimator is defined in a similar way.

4 Related nonparametric methods in reserving

The reserving literature contains other suggestions for smoothing methods, and these can be related to the kernel density methods proposed in this paper. The aim of this section is also to demonstrate the novelty of the approach in this paper.

An early paper where the notion of smoothing is applied to the reserving problem is Verrall (1996) which was followed by England and Verrall (2001). The latter paper formalizes a traditional approach in actuarial practice which consists of smoothing the development factors in the deterministic chain ladder approach. England and Verrall (2002) contains a description of this and further develops it in the framework of the generalized linear models (GLM). Chain ladder models are parametric models where the number of parameters increases in a linear way with the dimension of the run-off triangle. From a triangle of dimension m and assuming the Poisson model for the entries in the triangle with multiplicative structure such as (1), the log-likelihood can be written as

$$\mathcal{L}\left(\alpha_i, \beta_j; \aleph_m\right) = \sum \sum_{(i,j)\in\mathcal{I}_m} \left\{-\alpha_i\beta_j + \alpha_i\beta_j \log(N_{ij})\right\}, \tag{10}$$

(omitting a constant term). Verrall (1991) and Kuang, Nielsen and Nielsen (2009) proved that the maximum likelihood estimators, $\{\widehat{\alpha}_i, \beta_j, 1 \leq i, j \leq m\}$, provide the chain ladder estimates by $\widehat{N}_{ij} = \widehat{\alpha}_i\widehat{\beta}_j$, for each $i, j = 1, \ldots, m$. The intuition in this paper indicates that low levels of aggregations in the triangle would lead to serious problems in the likelihood behavior.

Apart from the well-known problem of identification of the parameters, which was solved by Kuang, Nielsen and Nielsen (2008), maximum likelihood methods tend to break down when the dimension goes to infinity. Grenander's method of sieves was suggested (Greman and Hwang 1982) as a method for modifying classical estimators so as to make them appropriate for such nonparametric problems. When the dimension of the parameter space goes to infinity, it is suggested that the optimization is attempted within a subset of the parameter space, with this subset then being alowed to increase with the sample size. The sequence of subsets from which the estimator is drawn is called a "sieve" and therefore the resulting estimation method is called the "method of sieves". Also the growth rate is controlled by the sieve size which, in practice, has to be chosen .

to illustrate the approach, we consider the example used by Greman and Hwang (1982) which is the histogram. For any univariate density f the maximum likelihood estimator based on an i.i.d. sample, $X_1, \ldots, X_n$, is defined as the maximizer of $\mathcal{L}(f; X_1, \ldots, X_n) = \Pi_{i=}^n f(X_i)$. If nothing is known about the target density the maximum cannot be achieved. This problem could be solved by restricting the set of candidates where the optimization is carried out. A sieve can be defined for this problem as the following sequence

$$s_{f,m} = \{f : f \text{ is a p.d.f. constant on } B_j \; j = 1, \ldots, m\}$$

with B_j being m contiguous intervals or bins dividing the support of f with bin size equal to Λ_m. Then the method of sieves consists of maximizing the likelihood $\mathcal{L}$ in the subspace $s_{f,m}$, allowing m to grow with the sample size n. The solution, which is thus a sieve estimator, is the well-known histogram estimator which was described in Section 2.1. Other examples in density estimation include penalized maximum likelihood estimators. A kernel smoother such as the simple Parzen-Rosemblat-type estimator, described in Section 3.1, is close to the sieves method but differs in the fact that it is not a maximum likelihood estimator for the density (see Greman and Hwang, 1982, for more details). Thus, the kernel approach in this paper moves away from the classical method of sieves to the modern and powerful kernel smoothing theory where good estimators such as the local linear and the multiplicative bias corrected estimators can attain good theoretical properties with excellent practical performance.

The flexible method proposed by England and Verrall (2001) is quite close to the above method of sieves. In this case the relation comes from an application to regression, also described by Greman and Hwang (1982), where a particular choice of sieve leads to the well-known smoothing splines estimator for the univariate regression function. Again, it is known that the maximum likelihood estimator is not consistent when a nonparametric (of infinite dimensional parametric) regression function is assumed. A sieve for this problem can be defined as the set of absolutely continuous functions r satisfying $\int |r'(x)|^2 \, dx \leq m$, and the sieves estimator would be a first degree polynomial smoothing spline (see for example Fan and Gijbels 1996 for a description of smoothing regression splines). England and Verrall (2001) exploits the regression formulation of the reserving problem to apply smoothing ideas. The approach of this paper becomes more complex than in Greman and Hwang (1982), but the aim is still to provide estimators having the maximum likelihood properties in a nonparametric framework. The aproach of England and Verrall (2001) is described here, in order to make the connection with the classical method of sieves clear, and also to show thel connection with the kernel methods suggested in this paper.

Consider the GLM formulation of the reserving problem described in Section 2, and assume a more general framework where the error distribution could be Poisson (as considered here) but also gamma, inverse gamma etc. Such a variety of distributions allows the GLMs to be used for bothe claim claims and claim amounts triangles, with the Poisson model being suitable for the counts triangle, and an overdispersion property often used for the triangle of paid claims. Denoting by C_{ij} the entries in any of these triangles the above distribution can be characterized by the following assumption for the variance:

$$\mathrm{V}(C_{ij}) = \phi m_{ij}^{\rho},$$

with $\mathrm{E}[C_{ij}] = m_{ij}$. Values of $\rho = 1, 2$ and 3 give the Poisson, gamma and inverse gamma, respectively. We focus on one of the GLM modesls for the mean for the data in a triangle, given by:

$$\log(m_{ij}) = c + a_i + b_j. \tag{11}$$

Here the use of the logarithmic transform makes the model linear but it also imposes some positivity constraints on the data in the triangle. This parametric model is usually fitted in practice using standard GLM tools. Note that if the dimension of the triangle is allowed to grow (and hence also the number of parameters in (11) increases), it would again give rise to a non-parametric maximum likelihood problem. Here the method of sieves would suggest defining a sieve candidate which could lead to consistent solutions. This was not the motivation of England and Verrall (2001), where a new and flexible model is described which could generalize simpler models such as that specified in equation (11) or others such as the so called Hoerl curve given by

$$\log(m_{ij}) = c + a_i + b_i \log(j) + \gamma_i j. \tag{12}$$

In this case, generalized additive models (GAMs) can be used of the following form:

$$\log(m_{ij}) = s_{\theta_i}(i) + s_{\theta_j}(\log(j)) + s_{\theta_j}(j). \tag{13}$$

In this, s_{θ_i} and s_{θ_j} are smoothers for the underwriting period i and the development period j, with smoothing parameters θ_i, and θ_j, respectively. The extremes values for the smoothing parameters i.e. zero and infinity, would lead to either the classical chain ladder model in (11) or the Hoerl curve in (12), respectively. Smoothing splines smoothers were considered for s_{θ_i} and s_{θ_j}, thereby considering the underwriting and development periods as continuous variables.

The relatively complex computational requirements of the GAMs in England and Verrall (2001) may have deterred them from becoming popular in reserving practice. A different approach, using GLMs was used by Björkwall, Hössjer, Ohlsson and Verrall (2011), where smoothing is introduced motivated by what is often done inl practice in reserving by an actuary: the smoothing of the development parameters using just ad hoc truncations of the estimated values from the chain ladder method.

All these smoothing approaches in reserving have a relationship with the general method of sieves, even though the nonparametric nature of the problem was not the original motivation. Also, none of these papers considered data recorded more frequently (or continuously) and they do not therefore address the nonparametric maximum likelihood problem.

Another application of smoothing ideas can be found in Verrall (1996), which suggested a simple smoothing of the chain ladder underwriting period parameters derived from the simple model in (11). The limited amount of data to estimate these underwriting period effects can lead to very volatile estimates, in which case the introdutione an appropriate level of smoothing can more stable and reliable results. Also having a similar aim but considering also smoothing in the development period are Zehnwirth (1989) and Verrall (1989), which use the Kalman filter.

The next section compares some of these approaches with the nonparametric density approach using real insurance data. But before to considering numerical results, where the peculiarities of the data might obscure some of the weakness of the methods being compared, we conclude this section with some remarks about the differences between previous approaches to introduce smoothing in reserving from that proposal in this paper.

From the above description, it is clear that the development of smoothing methods in reserving has focused on solving some of the drawbacks in traditional reserving practice such as the high volatility of results from classical chain ladder methods. Reducing volatility is indeed one of the results of using smoothing methods, but previous papers did not consider the implications of the nonparametric reserving problem. England and Verrall (2001) pointed out the continuous nature of the problem but it was simply noted that a continuous model was being applied to a nonparametric problem. Indeed even the continuous nature of the problem is recognised, all of the previous approaches look at the regression problem more than the actual density problem. The regression view of the density problem is indeed quite useful and has allowed many of the contributions to the density estimation problem to be developed. However, such a perspective relies on the histogram and therefore all the attempts to introduce smoothing methods would lead to smoothed versions of the histogram. There are many papers in the literature demonstrating that it is a poor and inefficient solution for the problem (see, for example, Jones, 1989). The new formulation of the problem as a density estimation problem therefore has wider implications and has not been considered before in the actuarial literature. This wider perspective allows us to develop modern and powerful nonparametric methods, which arel known to provide excellent results in other fields. This, although the proposal in paper may appear to be too sophisticated for stochastic reserving, we believe that it is in fact simpler and more intuitive.

5 Illustration with real insurance data

5.1 The data and classical chain ladder

In this paper we consider an illustration of the reserving problem trough a personal accident data set from a major insurer. These data were previously used by Martínez-Miranda, Nielsen and Verrall (2012) to estimate the reserve using incurred counts and paid data through a micro-model on the underlying individual data. In this study we restrict our attention to the incurred counts triangle. As we discussed along the paper a proper analysis of the paid data using the kernel density approach should require further work to deal properly with the correlations in the individual payments series.

The data set consists of quarterly data arranged into an incremental run-off triangle. Cells in the triangle correspond to number of reported counts as were formulated above in $\aleph_{m'} = \{N_{ij}; (i,j)\mathcal{I}_{m'}\}$, with $m' = 76$. Working with such level of aggregation is not recommendable when applying classical methods based on maximum likelihood (or quasi-likelihood) such as classical CLM. We remind to the reader our discussion in Section 4 where it was pointed out that maximum likelihood tends to break down when the estimation problem becomes actually a nonparametric problem. As we discussed in this section there are some approaches in the literature to deal with this issue without moving from the classical perspective. However the common method used for many years in practice consisted of going to higher levels of aggregation such as years so the dimension of the parametric space would be reduced. Let denote by $\aleph_m = \{N_{ij}; (i,j)\mathcal{I}_m\}$ (with $m = 19$) the yearly aggregation of the data which is plotted in Figure 1. The rows correspond to the underwriting years and the columns correspond to the delay until reporting, also in years from the underwriting year. Then the observations consist of a histogram with bins containing four quarters. Such a histogram is the starting point of traditional methods such as the classical chain ladder. The projections using this method are calculated assuming the mean model $\mathrm{E}[N_{ij}] = \alpha_i\beta_j$ $(i,j = 1,\ldots,m$ and $m = 19)$ and the parameters are estimated using maximum likelihood under the Poisson distribution. The resulting projections are shown in the bottom graph in Figure 1. We can see that these projections represent quite well the data histogram.

5.2 The continuous approach

Now we compare the classical chain ladder solution given above with the continuous density approach suggested in this paper. Since the lower level of aggregation in which the data are available are quarters we make use of discrete approximations of our local linear (LL) and multiplicative bias corrected (MBC) estimators. Such expressions are given in the Appendix A and use as input data the quarterly counts triangle with dimension $m' = 76$. To derive both LL and MBC estimators we should start by estimating an unstructured density in the observation set (in this case it is $\mathcal{I}_{m'}$). Such estimators are formulated in (22) and (23) for the available data and require a proper bandwidth choice. The crossvalidation method in (9) provides a suitable choice for the bandwidths which becomes $\widehat{h}_{cv} = (10.2, 2.9)$ for the LL estimator, and $\widehat{h}_{cv} = (10, 3.3)$ for MBC estimator. Note that it is requires to oversmooth in the underwriting direction but undersmooth in the development component. These unstructured estimators are the starting point of our density estimator for forecasting purposes. They have been plotted in the left panels of Figure 2. The estimated densities in the underwriting and development directions (f_1 and f_2) are shown in Figure 3 and compared with classical approaches such as classical chain ladder and two smoothing methods suggested in the classical reserving framework. These methods were described in Section 4 and we have focus in this application in two of them: a sieves method to smoothing the chain ladder parameters estimated from quarterly data (Verrall 1996 (check??)) and the following generalized additive model (GAM) suggested by England and Verrall 1991:

$$\log\left(\mathrm{E}[N_{ij}]\right) = s_{\theta_i}(i) + s_{\theta_j}(j), \tag{14}$$

with s_{θ_i} and s_{θ_j} being smoothing splines with crossvalidated smoothing parameter. The resulting components are shown in Figure 3 and compared with the derived using the density approach and classical chain ladder.

5.3 Prediction of the outstanding claims

In reserving the interest is often to provide a summary of the predicted outstanding claims more than individual estimates for cells in the input triangles or even for individual claims. In fact the reserving practice focuses in deriving predictions for the overall total (or the reserve) or totals for each of the future calendar year. From the classical CLM approach the outstanding numbers are obtained just by summing the predicted values for the claims in the lower triangle. Thus the predicted reserves for the future calendar period are derived just by summing up the diagonals in the lower triangle i.e.

$$\widehat{D}_k = \sum_{i=2}^{m} \widehat{N}_{i,m+k-i+1} = \sum_{i=2}^{m} \widehat{\alpha}_i \widehat{\beta}_{m+k-i+1} \tag{15}$$

for $k = 1, \ldots, m-1$. Thus the overall total will be $R = \sum_{t=1}^{m-q} D_t$. Under the continuous approach these predictions are defined from the following integrals:

$$\widehat{D}_t = \tau \int_0^T \widehat{f}_1(x)\widehat{f}_2(T - x + t)dx, \tag{16}$$

for calendar time $t+T$ with $t \in (0, T)$. Here τ is the total exposure in $\mathcal{I}$. The predictions derived from the classical CLM and the continuous approach (LL and MBC) are reported in Table 1. The results are compared also with the classical chain ladder projections and the two smoothing methods in classical reserving defined above.

We can see that the compared method are quite different so the natural question is to perform a validation of the methods for the analyzed data set. This is our aim in the next subsection but before moving there just to point out that the performance of the methods should be asses for different prediction goals. This is the best method to predict individual data (in this case cells in the yearly triangle) could not be the best when the aim is to predict an aggregation of the claims such as calendars or the overall total. In fact, using the smoothing ideas in this paper to predict cells should be required an undersmooth degree but the opposite will be required when the aim is to predict a total. Therefore we expect that smoothing methods works better to predict calendar years and total number of claims.

5.4 Validation

A common method used in reserving to validate the methods consists of testing against the experience. This is done through the so called backtest. The idea is quite simple: since we can only check the predictions with what we have already observed, then we simply reduce the data to estimate and use only the older data to predict the more recent data. Note that this process uses the key assumption that the past is a good predictor of the future.

Thus from the available data we have cut off a number of calendar periods, this is diagonals in the triangle, and keep the numbers to test later with the predictions. By denoting by c the number of cut periods, the dimension of the reduced triangle would be $m-c$ and the observation set $\mathcal{I}_c = \{(i,j); i,j =$

$1, \ldots, m-c, i+j-1 \leq m-c\}$. Now we can project from this reduced triangle in the future which is given by the set $\mathcal{J}_c = \{(i,j); i = 2, \ldots, m-c, j = m-c-i+1, \ldots, m-c\}$. And finally we compare the projections with the original kept data which spread out in $\widetilde{\mathcal{J}}_c = \{(i,j) \in \mathcal{J}_c; i+j-1 \leq m\}$.

We use a number of different measures for the error depending on which is the objective of prediction. Thus we are interested in validating the methods to achieve three possible aims:

1. To predict individual cells i.e. number of claims which incurred in the year i and will be reported with $j-1$ years of delay, N_{ij}.

2. To predict cash flows i.e. total number of claims which will be reported in the calendar year $t = i+j-1$, this is $D_{t;c} = \sum_{i,j:i+j-1=t,(i,j)\in\widetilde{\mathcal{J}}_c} N_{ij}$.

3. And to predict the overall total of claims in the future $R_c = \sum_{(i,j)\in\widetilde{\mathcal{J}}_c} N_{ij}$.

Therefore the performance of the methods applied in previous section to the data set should be evaluated in three different ways, depending on the prediction goal defined above.

$$\text{Cells:} \quad Rerr_1^c = \frac{\sum_{(i,j)\in\widetilde{\mathcal{J}}_c} (\widehat{N}_{ij} - N_{ij})^2}{(i,j) \in \widetilde{\mathcal{J}}_c N_{ij}^2} \tag{17}$$

$$\text{Calendar:} \quad Rerr_2^c = \frac{\sum_{t=1}^{c} (\widehat{D}_{t;c} - D_{t;c})^2}{\sum_{t=1}^{c} (D_{t;c})^2} \tag{18}$$

$$\text{Total:} \quad Rerr_3^c = \frac{|\widehat{R}_c - R_c|}{R_c} \tag{19}$$

The results for the backtest projecting from the reduced triangles $\mathcal{J}_c$ with $c = 1, 2, 3, 4$ and 5 years are reported in Table 2. The results from the test are quite unstable and do not provide a clear picture of which method is working better for the problem. In the next section we perform also a simulation study to provide clearer conclusions.

6 Simulation study

To do... We will simulate monthly data.

7 Conclusions

To write...

References

Bjïrkwall, S., Hïssjer, O., Ohlsson, E. and Verral, R.J. (2011), A generalized linear model with smoothing effects for claims reserving, *Insurance: Mathematics and Economics*, 49, 27–37.

England, P.D. and Verrall, R.J. (2001), A flexible framework for stochastic claims reserving, *Proceedings of the Casualty Actuarial Society* LXXXVIII, 1–38.

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

Geman, S. and Hwang, C-R. (1982), Nonparametric maximum likelihood estimation by the method of sieves, *Annals of Statistics*, 10(2), 401–414.

Jones, M.C. (1989) Discretized and Interpolated Kernel Density Estimates, *Journal of the American Statistical Association*, 84 (407) 733–741

Linton, O. and Nielsen, J.P. (1998) An optimization interpretation of integration and backfitting estimators for separable nonparametric models, *Journal of Royal Statistical Society*, Series B 60, 217–222.

Linton, O. and Nielsen, J.P. (1995) A kernel method of estimating structured nonparametric regression, *Biometrika*, 82, 93–100.

Kuang, D., Nielsen B. and Nielsen J.P. (2009) chain ladder as Maximum Likelihood Revisited. *Annals of Actuarial Science* 4, 105–121.

Mammen,E., Linton, O. and Nielsen, J.P. (1999) The existence and asymptotic properties of a backfitting algorithm under weak conditions, *Annals of Statistics*, 27, 1443–1490.

Marx, B.D. and Eilers, P.H.C. (1998) Direct generalized additive modeling with penalized likelihood, *Computational Statistics & Data Analysis*, 28, 193–209.

Nielsen, J.P. (1999) Multivariate Boundary Kernels from Local Linear Estimation, *Scandinavian Actuarial Journal*, 1, 93–95.

Norberg, R., (1986): A contribution to modelling of IBNR claims. *Scandinavian Actuarial Journal*, 155-203.

R Development Core Team (2006). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.

Verrall,R.J. (1989), A state space representation of the chain ladder linear model, *Journal of the Institute of Actuaries*, 116, 589–609.

Verrall, R.J. (1991), Chain ladder as Maximum Likelihood, *Journal of the Institute of Actuaries*, 118, 489–499.

Verrall, R.J. (1996), Claims reserving and generalized additive models, *Insurance: Mathematics and Economics*, 19, 31–43.

Verrall, R.J. (2004), A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claims reserving, *North American Actuarial Journal*, 8(3), 67–89.

Wüthrich, M. and Merz, M. (2008). Stochastic Claims Reserving Methods in Insurance. *Wiley.*

Yu, K., Park, B.U. and Mammen, E. (2008) Smooth backfitting in generalized additive models, *Annals of Statistics*, 36, 228–260.

Zehnwirth, B. (1989), Regression methods -applications. In: *Conference Paper. 1989, Casual Loss Reserve Seminar.* Available online at:

`http://www.casact.org/pubs/CLRSTrans/1989/825pdf.`

A Smoothing the multiplicative density from aggregated data

As we have described above the reserving problem consists of estimating a two-dimensional density with a support into a triangle. As was described in Section 2.1 the density problem estimation can be viewed as a regression problem on using binned data (Fan and Gijbels, 1996). Since data in insurance are usually presented in an aggregated way such as quarters, years or in the lower levels by months or even weeks this approach can be required by the available data. In this practical situation it is useful to rewrite the methods we introduce in this paper using regression formulation on the available aggregated data. The approach is then closely related to marginal regression method by Linton and Nielsen (1998). In this work the aim is simply minimizing the distance from a multivariate regression function to two multiplicative components. Also the subsequent multiplicative bias correction we also suggest for the density problem can be reformulated in terms of the nonparametric regression model by Linton and Nielsen (1994).

The regression formulation becomes more intuitive and at the moment more popular in the current reserving practice as we discussed in Section 4. We think that this section will be a bridge to connect with the classical reserving audience and at the same time a simple statistics exercise consisting in adapting continuous methods to practical problems where the data are given at some level of aggregation.

Let consider here the reserving problem to be solved from aggregated data into a run-off triangle such as the counts triangle, $\aleph_m$. As we describe in Section 2.1 The regression model for the underlying problem can be written by

$$N_{ij} = r(z_{ij}) + \varepsilon_{ij} \tag{20}$$

with $z_{ij} = (x_i, y_j)$ the points in the grid, for $(i,j) \in \mathcal{I}_m$. In this case the local linear estimator of the unstructured density resulting from solving the problem (5), for any given point $z_0 = (x_0, y_0)$ can be derived from the close minimization regression problem:

$$\begin{pmatrix} \widehat{\Psi}_0 \\ \widehat{\Psi}_1 \end{pmatrix} = \arg\min \sum_{(i,j)\in\mathcal{I}_m} \left[N_{ij} - \widehat{\Psi}_0 - \widehat{\Psi}_{11}(x_0 - x_i) - \widehat{\Psi}_{12}(y_0 - y_j) \right]^2 \mathcal{K}_h(z_{ij} - z_0) dz. \tag{21}$$

The solution $\widehat{\Psi}_0$ gives an estimator for $r(z_0)$. By denoting as $\widetilde{r}(z_0)$ such estimate, the density $f(z_0)$ can be estimated by the discrete approximation

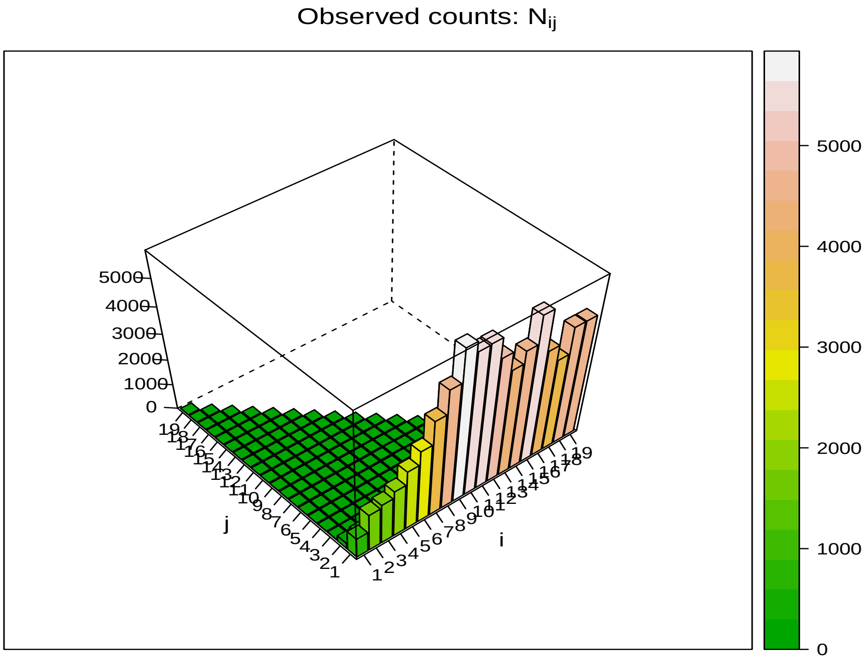
$$\widetilde{f}(z_0) = \widetilde{r}(z_0)/n\Lambda_m^2, \tag{22}$$

with $n = \sum_{(i,j)\in\mathcal{I}_m} N_{ij}$ and Λ_m the grid length. From arguments given in Section 3.1 we can see that the estimator in (22) is equal to the local linear density estimator in (5) as the grid-length Λ_m goes to 0. To derive the corresponding discrete approximation form the structured (multiplicative) density we carry out the the two-step method formulated in Section 3.3 from the just derived estimator (22).

Similarly we can derive the second improved smoother using bias reduction techniques that was proposed in Section 3.4. Again from the above regression view we can define the multiplicative bias correction estimator as was introduced by Linton and Nielsen (1994) for nonparametric regression. The unstructured MBC estimator at any point z_0 is defined from the local linear regression estimator $\widetilde{r}(z_0)$ by:

$$\widetilde{r}_{MBC}(z_0) = \widetilde{r}(z_0)\widetilde{h}(z_0) \tag{23}$$

where $\widetilde{h}$ is the local linear regression estimator calculated from a problem like (21) but using as responses the estimates $\{N_{ij}/\widetilde{r}(z_{ij}), (i,j) \in \mathcal{I}_m\}$. From the MBC estimator $\widetilde{r}_{MBC}$ the discrete unstructured MBC density is given by $\widetilde{f}_{MBC}(z_0) = \widetilde{r}_{MBC}(z_0)/n\Lambda_m^2$. And finally we use the two-step method to provide the discretized structured MBC estimator.



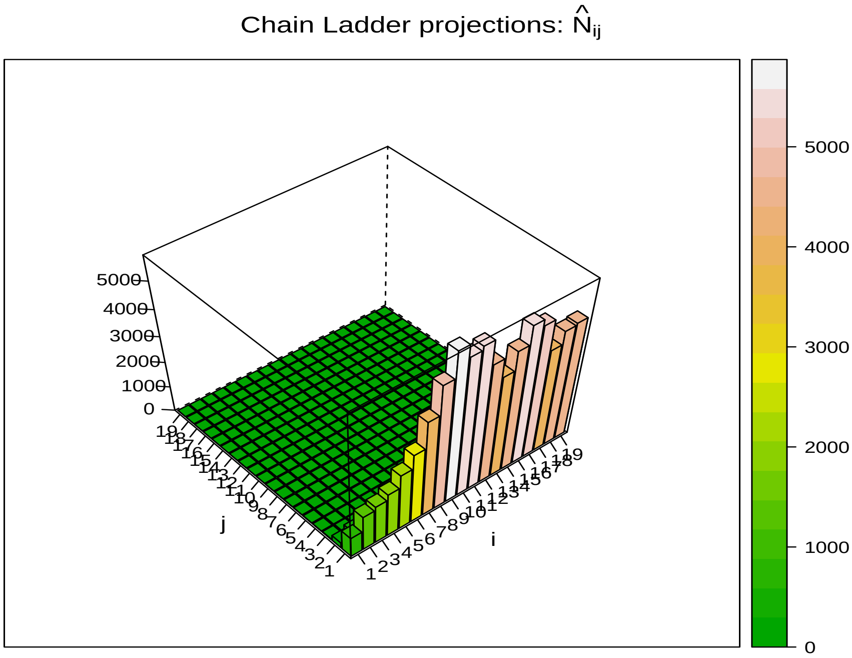


Figure 1: Insurance motor data: observed counts (run-off triangle) and classical chain ladder projections. The observed data consists of the number of claims which incurred at year i and were reported with a delay of $j-1$ years ($i, j = 1, \ldots, m$, $m = 19$). The bottom histogram shows the projections calculated using the chain ladder method.

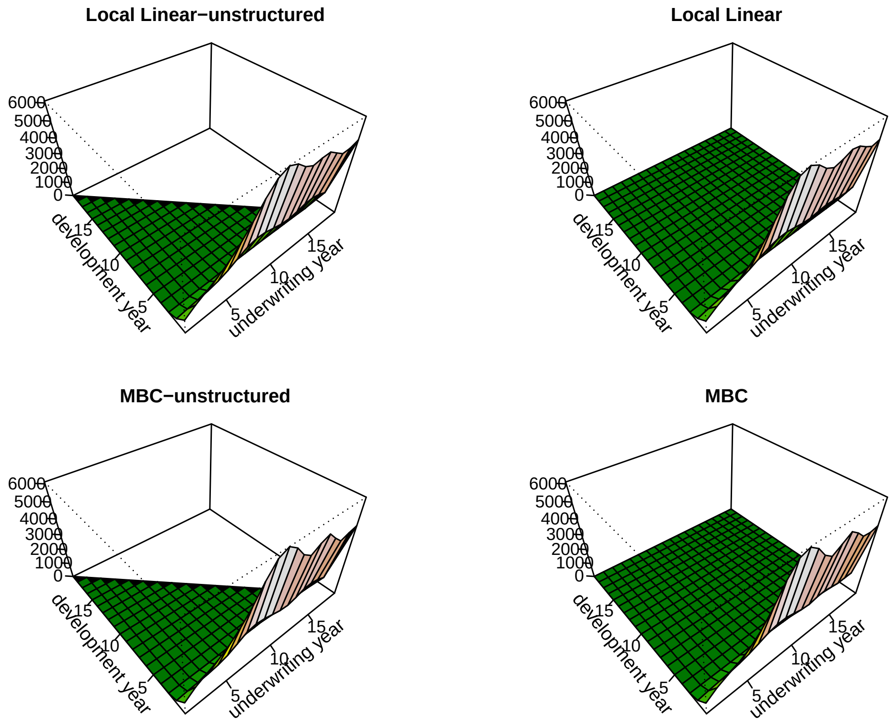


Figure 2: Forecasts for motor data using the continuous density approach. Left panels show the unstructured local linear (LL) and multiplicative bias corrected (MBC) estimators. The bandwidth parameters were chosen using crossvalidation that provided values $(h_1, h_2) = (10.2, 2.9)$ for LL and $(h_1, h_2) = (10, 3.3)$ for MBC. The estimated densities have been evaluated at years and the predicted number of claims have been plotted in the vertical axis. Right panels show the derived projections using the structured (multiplicative) estimators through the two step method in subsection 3.3.

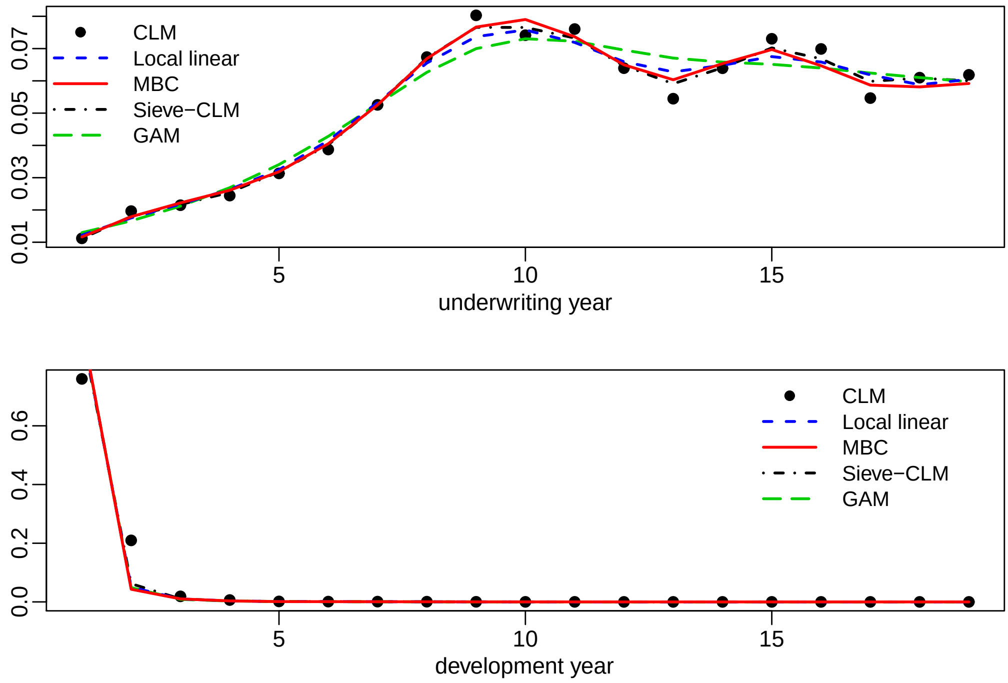


Figure 3: Estimated underwriting and development densities for motor data: incurred counts triangle. Top panel shows the resulting underwriting density (f_1) from LL and MBC. Similarly for the development density (f_2) in the bottom panel. The LL and MBC estimates are compared with the classical chain ladder parameters (CLM) and two smoothing related methods: a sieves method on quarterly chain ladder parameters (sieves-CLM) and a the GAM approach proposed by England and Verrall (1991).

Future	LL	MBC	CLM	Sieve-CLM	GAM
1	1647	1253	1425	1805	1239
2	187	161	181	200	151
3	68	60	69	69	55
4	32	26	30	32	28
5	21	17	20	21	17
6	14	12	15	14	11
7	9	7	9	9	8
8	5	4	5	5	5
9	3	2	3	3	3
10	2	2	2	2	2
11	1	1	1	1	1
12	1	1	1	1	1
13	1	0	1	1	1
14	1	0	1	1	0
15	1	0	0	1	0
16	0	0	0	0	0
17	0	0	0	0	0
18	0	0	0	0	0
Total	1992	1547	1763	2166	1524

Table 1: Predicted number of reported claims at each future calendar year. The predictions labeled by LL and MBC have been calculated from the structured local linear and multiplicative bias corrected estimators, respectively, with bandwidth estimated using crossvalidation. The column labeled as CLM provides the classical chain ladder predictions. The fifth column shows the results from a method of sieves on chain ladder estimates and the last column the predictions using the GAM suggested by England and Verrall (2001).

Objective	c	LL	MBC	Siev-CLM	GAM
Cells	1	2.4585	0.2354	5.5215	0.6241
	2	0.8593	1.2875	0.6553	0.8603
	3	0.8490	1.2782	1.0107	0.9895
	4	2.8465	0.9192	2.9988	0.2609
	5	0.6219	1.2662	0.5442	2.0258
Calendar	1	0.3114	0.0572	4.7132	0.3533
	2	0.8602	1.3575	0.6211	0.9406
	3	0.8552	1.1465	1.0001	1.0251
	4	2.7530	0.9022	2.8625	0.2215
	5	0.1441	1.1828	0.0989	1.7305
Total	1	0.3114	0.0572	4.7132	0.3533
	2	0.8628	1.4226	0.5867	0.9710
	3	0.8563	1.1816	1.0095	1.0572
	4	2.3621	0.8920	2.4244	0.0818
	5	0.3264	1.2409	0.1470	1.9921

Table 2: Back-testing on real insurance motor data. The relative prediction error has been evaluated when the aim is either to predict individual cells in the yearly triangle, total quantity for future calendar years (diagonals) or the overall total. The columns labeled as LL and MBC show the reduction/increment in the prediction error against the classical CLM considering the structured local linear and multiplicative bias correction estimated densities, respectively. The exact expression of these relative measures is given in (17), (18) and (19). The numbers in the table corresponds to the ratio of the measures for each method and the obtained for CLM. Thus quantities lower than 1 indicate an improvement on CLM.