City Research Online

City, University of London Institutional Repository

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: http://openaccess.city.ac.uk/16953/

Link to published version: http://dx.doi.org/10.1080/03461238.2016.1240709

City Research Online: http://openaccess.city.ac.uk/   publications@city.ac.uk

On the relationship between classical chain ladder and granular reserving

M. Hiabu

Cass Business School, City University London
munir.hiabu.1@cass.city.ac.uk

November 15, 2016

We connect classical chain ladder to the continuous chain ladder model of Martínez-Miranda et al. (2013). This is done by defining explicitly how the classical run-off triangles are generated from *iid* observations in continuous time. One important result is that the development factors have a one to one correspondence to a histogram estimator of a hazard running in reversed development time. A second result is that chain ladder has a systematic bias if the row effect has not the same distribution when conditioned on any of the aggregated periods. This means that the chain ladder assumptions on one level of aggregation, say yearly, are different from the chain ladder assumptions when aggregated in quarters and the optimal level of aggregation is a classical bias variance trade-off depending on the data-set. We introduce 'smooth development factors' arising from non-parametric hazard kernel smoother improving the estimation significantly.

Keywords: Chain Ladder, Granular Reserving, Development Factors, Solvency II, Non-Life Insurance

1. Introduction

Reserving is the process behind setting capital reserves for outstanding liabilities in non-life insurance. Insurance companies are obligated to account for claims that have been reported but not settled yet and also for incurred claims which have not even been reported. The reserve is often the major part of a non-life insurers balance sheet. Accurate estimation is necessary for pricing future policies and also for the assessment of solvency and net worth of the company. This in turn plays a major role in decisions for financial investments and also for sales or acquisitions of insurances. Finally, wrong

assessment can lead to bankruptcy of major companies, with consequences for the whole economic system; for example in the UK, the non-life insurance market accounts for 5% of the gross national product. These considerations come in hand with a growing sense that the reserving process has to be done more rigorous including accurate point forecast and discussions about its uncertainty around.

In practice, actuaries usually use the chain ladder method to calculate the reserve. The method is based on historical data aggregated as run off triangles, i.e., paid claims, claims counts, or incurred claims. For the sake of simplicity of the mathematical arguments we only consider claim counts. Chain ladders development factors (see (7)) are hereby the central object. One expression of their importance is maybe its many names: CL (chain ladder) factor, link-ratio, age to age factor, or forward factor. But despite its central role and intuitive appeal, as of today, practitioners and also academics are struggling with the understanding of development factors in terms of classical mathematical statistics. This might have let the author in Schmidt (2012) saying:

> " […] loss reserving is an art of which statistics is, although important, just a part."

This goes in hand with England and Verrall (2002) remarking on the usual reserving practise that

> "very often, the chain ladder technique is the first method to be applied, followed by manual smoothing of the resultant development factors, then adjustment of the results in line with expert opinion combined with additional information".

With these statements in mind, the reserving task remains, by its very nature, a statistical problem. Hence, a better statistical understanding of those practices and the reserving problem is necessary not only to to get reasonable point estimates and to quantify the risk and uncertainty in a reproducible way but also for understanding the underlying assumptions under which these results hold.

A main result of this paper is that when the classical run-off triangles are modeled as arising from observations in continuous time, then there is a quite easy understanding of the development factors in terms of mathematical statistics. We will show that there is a one to one correspondence to a histogram estimator of a hazard function (also known as force of mortality in the actuarial branch of longevity) in reversed development time. In Section 3, we show that

$$\widehat{\lambda}_j = \{1 - \widehat{\alpha}^H(T - x_j)\}^{-1},$$

where $\widehat{\lambda}$ are the development factors defined in (7) and $\widehat{\alpha}^H$ is the histogram estimator of the hazard function, see (10). This translates the estimation problem of development factors to the well known estimation problem of a hazard function in survival analysis. In this survival analysis framework it is possible to relax classical assumptions of chain ladder, for instance by allowing calendar time effects. A possibility of extension is adding covariates when estimating the hazard. Both possibilities transform the one

dimensional hazard to the multivariate case. In this paper we improve chain ladder with a third possibility and make the maybe most easy improvement in estimating the development factors. We replace the histogram estimator of the hazard by more efficient non-parametric kernel smoother of the hazard, see also Hiabu et al. (2016). The one to one correspondence then leads to non-parametric kernel smoothed development factors.

Modeling the complete data generating process leads to another discovery in this paper. An underlying assumption of any stochastic model describing the classical chain ladder method is the independence of underwriting date (row) effect and delay (column) effect, since the development factors do not depend on the underwriting date. In Proposition 3.1 below, it is shown that if this holds on the individual level, then chain ladder in its aggregated form is only consistent if the underwriting effect is identically distributed within a period. Already the simple example of a continuously linear increasing trend in the book-size will make chain ladder in-consistent by adding a systematic bias; more precisely the reserve will be overestimated in that case. Hence, if one does not see the aggregation in classical chain ladder as a smoothing step where the aggregation level converges to zero with growing sample size, then this should indeed be seen as the underlying assumptions of chain ladder.

There has been a lot of literature aimed at building a statistical model around the chain ladder method; see Kremer (1982), Verrall (1991), Mack (1993), Renshaw and Verrall (1998), Wüthrich, Merz, and Bühlmann (2008), and Kuang, Nielsen, and Nielsen (2009) among others. In the last years there is a growing sense in the industry that aggregated data or macro data is not accurate enough and maybe outdated in times of big data. This argumentation is not completely correct, since it is the very aim of statistics to compress information into a single number or function. Therefore, aggregation should be seen as a statistical pre-smoothing step. The problem then, however, is that there is a) little discussion about the optimal level of aggregation, which of course varies with the data at hand, and b) no discussion about the underlying individual data which justifies this type of aggregation. While discussing the underlying model of chain ladder, the papers mentioned before do especially not discuss the data generating process or sampling scheme which make it hard to understand and justify the implicit assumptions. Renshaw and Verrall (1998), for instance, say that the chain ladder method assumes "stationarity of the reporting process" without further defining what this process is and how the actual data arises from this process. Finally, granular methods are necessary if one is interested in a more detailed cashflow.

Models on individual data and continuous time have been developed by Arjas (1989) and Norberg (1993), where the individual claim development is modeled as a marked point process. These more theoretical contributions have been made more applicable through the work of Antonio and Plat (2014). A different semi-parametric approach based on copulas is given in Zhao, Zhou, and Wang (2009) and Zhao and Zhou (2010). A comparison of an individual model and chain ladder estimates derived from its aggregation are discussed in Huang, Wu, and Zhou (2016).

In recent literature there have also been developed models based on individual data assumptions but where the traditional run-off triangle data structure from chain ladder is kept. The idea is to keep the triangular structure and thus do not completely

throw away existing reserving theory and practice. Verrall, Nielsen, and Jessen (2010), Martínez-Miranda, Nielsen, and Verrall (2012), Hiabu et al. (2015), and Schiegl (2015) have assumptions on the individual data but work entirely with aggregated observations. This makes it hard to check the underlying assumptions on the individual data. Drieskens et al. (2012), Rosenlund (2012), Pigeon, Antonio, and Denuit (2013), and Godecharle and Antonio (2015) rely on individual data but work on aggregated time. Martínez-Miranda et al. (2013) formulated a continuous chain ladder model which keeps the traditional run-off triangle structure of classical chain ladder but considers individual data in continuous time.

This paper aims to connect and compare the continuous chain ladder model of Martínez-Miranda et al. (2013) and classical chain ladder. In practice one can imagine the continuous model being based on the classical data by defining a period as a second instead of, say, a year, which results in a triangle of only 0's and 1's. An important result of this paper is that chain ladder's estimation techniques corresponds to survival analysis techniques when the development time is reversed. With the time reversal one does not need exposure data to estimate the quantities of interest which is also the case in the classical chain ladder method. This is different to the individual data approaches based on Arjas (1989), Norberg (1993), and Antonio and Plat (2014), and will be explained in more detail in the next section.

2. The continuous model

2.1. Model formulation

In this chapter we will formulate the stochastic model of continuous chain ladder (Martínez-Miranda et al., 2013), and afterwards embed it into a counting process framework. For a better understanding in what follows it is helpful to be familiar with the classical chain ladder method, in particular the run-off triangle, see for example Taylor (1986) and Wüthrich and Merz (2008). The idea of continues chain ladder is that claims are point observations on the usual run-off triangle rather than being aggregated into bins as is assumed in the classical chain ladder method. With the counting process formulation we will then define a hazard function as part of the counting process intensity. In chapter 3, we will show that the hazard in reversed development time is the continuous version of the well known development factors of the chain ladder method.

The data in the classical chain ladder method is given as a run-off triangle and is one half of the square including the future claims which are needed to be estimated. We consider the probability space, $(\mathcal{S}, \mathcal{B}(\mathcal{S}), P)$, where $\mathcal{S}$ is the square $\{(x, y) : 0 \leq x, y \leq T\}$. The underwriting date, Y, and the reporting delay, X, of a claim are hence random with probability measure P, which describes how likely it is to see a claim on a certain position on the square. Since S is bounded by T, we implicitly assume that firstly all claims are reported within a maximum delay of T from their underwriting date and secondly that we have T time units of observed underwriting dates. Generally to avoid extrapolation, which we are not doing here, the maximum delay of a claim must be smaller than the range of observed underwriting dates. The theory described here would, as chain ladder

does, work if X and Y would be bounded by T_1, T_2, with $T_1 \leq T_2$. To simplify the notation, we have assumed $T = T_1 = T_2$. If this is not the case in practise, i.e $T_1 < T_2$, the remaining columns can just be filled with zeroes to obtain the same results.

We will assume that the density with respect to the Lebesgue measure, $f = \mathrm{d}P/\mathrm{d}\lambda$, is well defined and multiplicative, i.e., $f(x, y) = f_1(x)f_2(y)$. Hence, we assume that the components X and Y are independent. This assumption can be checked by usual independence tests, see Tsai (1990), Mandel and Betensky (2007), and Addona, Atherton, and Wolfson (2012). A more pragmatic solution is plotting the individual development factors and checking whether they lie on a horizontal line, see Hiabu et al. (2016).

We further assume that observations are only sampled on a subset of the full support of the density f. The truncated density is supported on the triangle, $\mathcal{I} = \{(x, y) : 0 \leq x, y \leq T, x + y \leq T\}$ - the well known run-off triangle. In this case, we consider observations of n independent and identically distributed claims, $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$, with $X_i \leq T - Y_i$, or equivalently $Y_i \leq T - X_i$, where T is the calendar time the data are collected. Note that (X_1, Y_1) is not distributed according to P and does not have density f, since we already know that it is on the upper triangle. Hence, its density is given by $f(x, y)/\int_{\mathcal{I}} (x, y)\mathrm{d}x\mathrm{d}y$. The observation schemes, $X_i \leq T - Y_i$, and $Y_i \leq T - X_i$ can be understood as random right-truncation when targeting only X or Y, respectively.

The well established method to make inference on such observation schemes is to reformulate the problem into a counting process framework, see for example Andersen et al. (1993). In the following we will focus on inference on the reporting delay X. Due to symmetry all the results can be easily adapted for the random variable Y. The development factors in chain ladder only correspond to estimates depending on X. To this end, we define a counting processes indicating the occurrences of X_i, $i = 1 \ldots, n$. A crucial point here is that right-truncation is not tractable as such, since the exposure is not observable: In the counting process model, one needs to know at every point in time how many individuals are at risk. Assume that we move T years forward in time and hence know about every claim on the square. Exposure in x is then the amount of claims having a greater reporting delay then x but could have been observed already at point of data collection if the delay would have been exactly x. This amount is not know at time of data collection, see Figure 1.

By reversing the time of the counting process, however, the right-truncation becomes a left-truncation, see for example Ware and DeMets (1976) and Lagakos, Barraj, and De Gruttola (1988), and exposure is observable, since all past claims are known. Note that in the models of Arjas (1989), Norberg (1993), and Antonio and Plat (2014) time is not reversed, and hence extra exposure data is needed to calibrate their model.

We define the time reversed counting processes as

$$N_i(t) = I\left(T - X_i \leq t\right), \quad (i = 1, \ldots, n),$$

where I denotes the indicator function, with respect to the filtration

$$\mathcal{F}_t^i = \sigma\left(\left\{(T - X_i) \leq s : \ s \leq t\right\} \cup \left\{(Y_i) \leq s : \ s \leq t\right\} \cup \mathcal{N}\right),$$
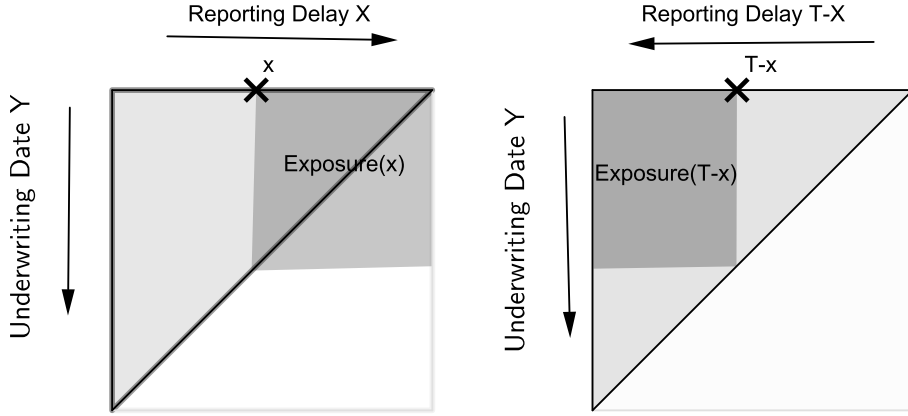
Figure 1: The exposure in forward moving time (left) and reversed time (right). Only in reversed time we observe the exposure.

satisfying the usual conditions, and where $\mathcal{N} = \{A : A \subseteq B, \text{with } B \in \mathcal{B}(\mathcal{S}), P(B) = 0\}$. Adding the null set, $\mathcal{N}$, to the filtration guarantees its completeness. This is a technically useful construction, but is not strictly necessary, since the subsequent results also hold if one does not assume completeness of the filtration, see Jacod (1979) and Jacod and Shiryaev (1987).

The random intensity of N_i, ν_i, is well-defined since X is absolutely continuous. It can be described, almost surely, through $\nu_i(t) = \lim_{h\downarrow 0} h^{-1} E\left[N_i\{(t+h)-\} - N_i(t-)\mid \mathcal{F}_{t-}^i\right]$. Straightforward computations lead to Aalen's multiplicative intensity model (Aalen, 1978):

$$\nu_i(t) = \alpha(t) Z_i(t),$$

where the hazard ratio α, and the predictable filtering process (individual exposure), Z_i, are

$$\alpha(t) = \lim_{h\downarrow 0} h^{-1} pr\left\{(T-X) \in [t, t+h) \mid (T-X) \geq t\right\} = \frac{f_1(T-t)}{F_1(T-t)} = \frac{f_1^R(t)}{S_1^R(t)},$$

$$Z_i(t) = I\left\{Y_i < t \leq (T - X_i)\right\},$$

and $F_1 = \int_0^{\cdot} f_1(x) dx$ is the cumulative distribution function. The crucial point in Aalen's multiplicative intensity model is that the hazard function, α, does not depend on Y. In chapter 3 we will show that the hazard in reversed development time, α, is the continuous version of the well known development factors λ of the chain ladder method. Before finishing this chapter, we introduce the notation $N(t) = \sum N_i(t)$ and the exposure $Z(t) = \sum Z_i(t)$.

2.2. Estimation in the continuous framework

In this section we briefly introduce three nonparametric estimators of the hazard function α in the continuous time framework: The histogram estimator, the local constant estimator and the local linear estimator. The local linear and the local constant estimator are well studied in the statistical literature of kernel smoothing, and we will only state the results and properties of the estimator for people not familiar with smoothing theory. The histogram estimator is known from applied fields as in age-period-cohort models of demographic problems.

An alternative to estimate the hazard function α would be to assume a parametric form on the intensity ν_i, see Borgan (1984) and Andersen et al. (1993). We chose not to do so in this paper, since a nonparametric estimation technique is more in the spirit of the chain ladder technique.

For the asymptotic properties we consider the following assumptions.

Assumption (S)

S1. The bandwidth $h = h(n)$ satisfies $h \to 0$ and $nh \to \infty$ for $n \to \infty$.

S2. The hazard function α is strictly positive and it holds that $\alpha \in C_2([0,T])$.

S3. The kernel K is symmetric, has bounded support and has a second moment.

Assumptions (S1) - (S3) are standard regularity assumptions in smoothing theory (Silverman, 1986; Simonoff, 1998). Note that under assumption (S2), the asymptotic relative exposure $\gamma(t) = pr(Z_1(t) = 1)$ is continuous and from empirical process theory it is known that

$$\sup_{s \in [0,T]} |Z(s)/n - \gamma(s)| = o_p(1). \tag{1}$$

2.2.1. The histogram estimator of the hazard

The maybe simplest way to derive an estimator of the hazard function, α, is the histogram estimator. Let's assume that a parameter, $h > 0$, as bin width is given. A histogram estimator of α on equally sized bins, with bin size h is derived by dividing the number of observations - relative to the bin width - in one bin by the number of exposure at that bin. For t in the bin $[c1, c2)$, that is

$$\widehat{\alpha}_h^H(t) = \frac{h^{-1} \sum_{i=1}^{n} \int_{c1}^{c2} \mathrm{d}N_i(s)}{\sum_{i=1}^{n} \int_{c1}^{c2} Z_i(s)\mathrm{d}s} = \frac{O^H(t)}{E^H(t)}. \tag{2}$$

In Hoem (1969) optimality of the histogram estimator is proven if the true hazard, α, is constant on the bins. The following proposition shows the asymptotic properties when local constancy is not assumed. The proof can be found in the Appendix C.

Proposition 2.1 Assume that assumptions (S1)-(S2) are satisfied. The histogram estimator has asymptotically a bias of order h and variance of order $(nh)^{-1}$. More precisely, the following pointwise asymptotics holds for $t \in (0, T)$:

$$(nh)^{1/2} \left\{ \widehat{\alpha}_h^H(t) - \alpha(t) - B(t) \right\} \xrightarrow{\mathfrak{D}} N \left\{ 0, \sigma^2(t) \right\},$$

where

$$B(t) = \alpha'(t) h^{-1} \int_{c_1}^{c_2} (t - s) \, \mathrm{d}s + o(h), \quad \sigma^2(t) = \alpha(t) \gamma(c_2)^{-1}.$$

2.2.2. Local polynomial estimator of the hazard

The idea of local polynomial fitting is quite old and might originate from early time series analysis, see Macaulay (1931). It has been adapted to the regression case in Stone (1977) and Cleveland (1979). A general overview of local polynomial fitting can be found in Fan and Gijbels (1996). The local constant estimator has a reduced convergence rate at boundaries. This is not the case for polynomials of order $p \geq 1$. In general, a higher order reduces bias but increases variance. But variance only increases when the order changes from odd to even. In this paper we will only consider the cases $p = 0, 1$, that is the local constant and the local linear estimator of the hazard function.

We define the local constant estimator, $\widehat{\alpha}_{h,K}^{LC}(t)$ of $\alpha(t)$, as the minimizer, $\widehat{\Theta}_0$, in the equation

$$\widehat{\Theta}_0 = \arg \min_{\Theta_0 \in \mathbb{R}} \sum_{i=1}^{n} \left[\int K_h(t - s) \Theta_0^2 Z_i(s) \mathrm{d}s - 2 \int K_h(t - s) \Theta_0 \, \mathrm{d}N_i(s) \right], \qquad (3)$$

where for a given kernel, K, and a bandwidth, h, $K_h(t) = h^{-1} K(t/h)$. The definition of the local constant estimator as the minimizer of (3) can be motivated by the fact that its minimizer equals the least square criteria,

$$\arg \min_{\Theta_0 \in \mathbb{R}} \left(\lim_{\varepsilon \to 0} \sum_{i=1}^{n} \int \left[\left\{ \frac{1}{\varepsilon} \int_s^{s+\varepsilon} \mathrm{d}N_i(u) - \Theta_0 \right\}^2 - \xi(\varepsilon) \right] \times K_h(t - s) Z^i(s) \, \mathrm{d}s \right),$$

where $\xi(\varepsilon) = \{ \varepsilon^{-1} \int_{s-\varepsilon}^{s} \mathrm{d}N^i(u) \}^{-2}$ is a just a vertical shift which is added to make the expression well-defined. The solution of (3), see also Nielsen and Tanggaard (2001), is given by

$$\widehat{\alpha}_{h,K}^{LC}(t) = \frac{\sum_{i=1}^{n} \int K_h(t - s) \mathrm{d}N_i(s)}{\sum_{i=1}^{n} \int K_h(t - s) Z_i(s) \mathrm{d}s} = \frac{O^{LC}(t)}{E^{LC}(t)}. \qquad (4)$$

For every Kernel K we define

$$\mu_i(K) = \int s^i K(s) \mathrm{d}s, \quad R(K) = \int K^2(s) \mathrm{d}s.$$

The following proposition states that the local constant estimator is efficient in optimal rate sense.

Proposition 2.2 (Hjort, West, and Leurgans (1992)) Assume that assumption (S) is satisfied. Then, the following pointwise asymptotics holds for $t \in (0, T)$:

$$(nh)^{1/2} \left\{ \widehat{\alpha}_{h,K}^{LC}(t) - \alpha(t) - B(t) \right\} \xrightarrow{\mathfrak{D}} N \left\{ 0, \sigma^2(t) \right\},$$

where

$$B(t) = \mu_2(K)h^2 \left\{ \frac{1}{2}\alpha''(t) + \alpha'(t)\gamma'(t)\gamma(t)^{-1} \right\} + o(h^2), \quad \sigma^2(t) = R(K)\alpha(t)\gamma(t)^{-1}.$$

Note that this result only holds in the interior of the support $[0, T]$. Following the proof one can easily see that the bias is of order b in the boundary region, i.e., the intervals $[0, b)$ and $(T - b, T]$. There have been several estimators proposed to derive convergence in the full support. Due to its simplicity but also convincing properties the local linear estimator became the maybe most popular kernel smoother. Similarly to the local constant estimator, we define the local linear estimator (Nielsen, 1998), $\widehat{\alpha}_{h,K}^{LL}(t)$ of $\alpha(t)$, as the minimizer, $\widehat{\Theta}_0$, in the equation

$$\begin{pmatrix} \widehat{\Theta}_0 \\ \widehat{\Theta}_1 \end{pmatrix} = \arg\min_{\Theta_0, \Theta_1 \in \mathbb{R}} \sum_{i=1}^{n} \left[\int K_h(t - s) \left\{ \Theta_0 - \Theta_1(t - s) \right\}^2 Z_i(s)\mathrm{d}s \right.$$
$$\left. - 2 \int K_h(t - s) \left\{ \Theta_0 - \Theta_1(t - s) \right\} Z_i(s)\mathrm{d}N_i(s) \right]. \quad (5)$$

With solution

$$\begin{aligned} \widehat{\alpha}_{h,K}^{LL}(t) &= n^{-1} \sum_{i=1}^{n} \int \overline{K}_{t,h}(t - s)\mathrm{d}N_i(s) \\ &= \frac{\sum_{i=1}^{n} \int K_h(t - s) \left\{ a_2(t) - a_1(t)(t - s) \right\} \, \mathrm{d}N_i(s)}{\sum_{i=1}^{n} \int K_h(t - s) \left\{ a_2(t) - a_1(t)(t - s) \right\} Z_i(s) \, \mathrm{d}s} \\ &= \frac{O^{LL,\delta}(t)}{E^{LL,\delta}(t)}, \end{aligned} \quad (6)$$

where

$$\overline{K}_{t,h}(t - s) = \frac{a_2(t) - a_1(t)(t - s)}{a_0(t)a_2(t) - \{a_1(t)\}^2} K_h(t - s),$$

and

$$a_j(t) = n^{-1} \int K_h(t - s)(t - s)^j Z(s)\mathrm{d}s \quad (j = 0, 1, 2).$$

The notation of $\overline{K}_{t,h}$ is chosen because it is indeed, given (S3), a second order kernel with respect to the measure $Z(s)\mathrm{d}s$:

$$n^{-1} \int \overline{K}_{t,h}(t - s)Z(s)ds = 1, \quad n^{-1} \int \overline{K}_{t,h}(t - s)(t - s)Z(s)ds = 0,$$

9

$$n^{-1} \int \overline{K}_{t,h}(t-s)(t-s)^2 Z(s)ds > 0.$$

Furthermore, Nielsen and Tanggaard (2001) showed that $\overline{K}_{t,h}(t-s)$ is asymptotically equivalent to $k_{t,h}(t-s)Z^{-1}(s)$, with

$$k_{t,h}(t-s) = \frac{c_2(t) - c_1(t)(t-s)}{c_0(t)c_2(t) - \{c_1(t)\}^2} K_h(t-s), \quad c_j(t) = n^{-1} \int K_h(t-s)(t-s)^j ds,$$

which in turn pointwise equals $K(t-s)$, for n large enough. This considerations make it not surprising that the local linear estimator has similar point-wise asymptotics as the local constant estimator.

Proposition 2.3 (Nielsen, 1998) Assume that assumption (S) is satisfied. Then, the following asymptotics holds for $t \in (0, T)$:

$$(nh)^{1/2} \left\{ \widehat{\alpha}_{h,K}^{LL}(t) - \alpha(t) - B(t) \right\} \xrightarrow{\mathfrak{D}} N \left\{ 0, \sigma^2(t) \right\},$$

where

$$B(t) = \frac{1}{2}\mu_2(K)\alpha''(t)h^2 + o(h^2), \quad \sigma^2(t) = R(K)\alpha(t)\gamma(t)^{-1}.$$

3. Discretization of the continuous model

3.1. The model

In the previous chapter we have defined several estimators of the hazard function, given observations in continuous time. In the non-life insurance context, data are usually aggregated in so called yearly or quarterly run-off triangles. This is done by aggregating the continuous triangle, $\mathcal{I}$, into a grid of parallelograms, see Figure 2.

The observation scheme is very similar to those in a Lexis diagram known from age-period-cohort models in demography, and aggregation of the same parallelograms is there known as the first principle set (Hoem, 1969; Keiding, 1990). In the language of age-period-cohort models, the form of a parallelogram arises because while data is collected with respect to cohort (underwriting date) and year (claim delay), the aggregation is done with respect to cohort and period (calendar time). While aggregation into squares would make many things easier, the triangular observation scheme would then imply that the number of observation changes with different aggregation level, and in particular forecasting would not be possible for the last underwriting period.

Let δ be the grid width with integer valued inverse. The individual data of independent, identically, distributed data (X_i, Y_i) are aggregated to observations, (X_i^δ, Y_i^δ), with support on

$$\mathcal{I}^\delta = \{(x_j, y_k) = ((j+0.5)\delta, (k+0.5)\delta) : j, k = 0, 1 \ldots, T\delta^{-1} - 1, j + k \leq T\}.$$

The discrete observations are then described via

$$(X_i^\delta, Y_i^\delta) = (x_j, y_k) \Leftrightarrow Y_i \in [k\delta, (k+1)\delta) \text{ and } X_i + Y_i \in [(j+k)\delta, (j+k+1)\delta)$$
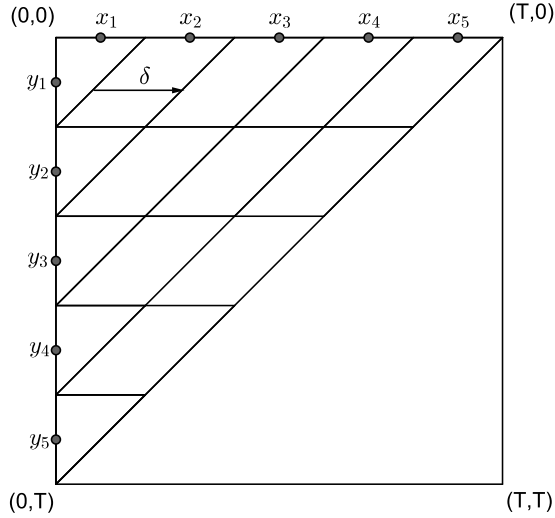
Figure 2: The usual aggregation of a triangle in the chain ladder method. The bin-width δ represents the length of a period.

Note that this implies that

$$X_i \in [(j-1)\delta \vee 0, (j+1)\delta).$$

The parallelogram aggregation adds a non-trivial dependency structure between the components X^δ and Y^δ, even though X and Y might be independent. The chain ladder method implicitly assumes independence between underwriting date and development delay, since the development factors do not change for different underwriting dates, see also various paper discussing the underlying model of chain ladder, e.g., Mack (1993) and Renshaw and Verrall (1998).

It is then important to note that the necessary independence of the components of (X^δ, Y^δ) does generally not follow from the independence of X and Y. Consider the following assumptions.

Assumption (D)

D(i) The density function f_2 of Y is multiplicatively separable in the sense that there exist functions g_1 and g_2 so that for every $y \in [y_j - 0.5\delta, y_j + 0.5\delta)$, it holds that $f_2(y) = g_1(y_j)g_2(y - y_j)$.

D(ii) The random variables X^δ and Y^δ are independent.

Assumption $D(i)$ basically says that the random variable Y needs to have the same distribution when conditioned on any bin on the grid.

11

Sufficient conditions are for instance local constancy, $f_2(y) = g(y_k)$, as most often assumed in age-period-cohort literature, or an exponential growth $f_2(y) = c\exp(y)$, where c is a norming constant.

Note that those assumptions can not be checked if one has only access to the aggregated data, also whether assumption D(i) or D(ii) are satisfied or by how much they violated does depend on the level of aggregation.

Proposition 3.1 If X and Y have a multiplicative separable density, i.e., $f(x,y) = f_1(x)f_2(y)$, and f_1 is not further specified, then assumption D(i) and D(ii) are equivalent.

The proof can be found in the Appendix C.

The implication of this proposition is that if assumption D(i) is not satisfied, then the level of aggregation in chain ladder is a bias-variance trade off. The optimal level of aggregation should then be derived via a cross-validation method which needs to be developed.

The classical run-off triangle data is given in the form $(N_{r,s})$, $r,s = 1,\ldots,T$, $r+s \leq T+1$, and are the total numbers of claims of insurance incurred in period (most often a year or quarter) r which have been reported in period $r+s$, i.e., with s periods delay from r. The triangle is derived from the random variables (X^δ, Y^δ), via

$$\mathcal{N}_{r,s} = \#\{i : Y_i^\delta = r, X_i^\delta = s\} = \sum_{i=1}^n I\{Y_i^\delta = r, X_i^\delta = s\}.$$

Assuming there is no so called tail, i.e., all claims are reported within T periods, forecasts are obtained by estimating and summing up the values $(N_{r,s})$, $r,s = 1,\ldots,T$, $r+s \geq T+2$. For this, chain ladder estimates development factors,

$$\widehat{\lambda}_s = \frac{\sum_{k=1}^{T-s+1}\sum_{l=1}^s N_{k,l}}{\sum_{k=1}^{T-s+1}\sum_{l=1}^{s-1} N_{k,l}}, \quad (s = 2,\ldots,T), \tag{7}$$

and forecasts are derived as $\widehat{N}_{r,s} = N_{r,T-r+1}\prod_{l=T-r+1}^s \lambda_l$.

For deriving estimators in the discrete framework, we want to use the already developed theory in the continuous time case and introduce f_1^δ as density of X^δ with respect to the counting measure $\mu(A) = \delta\#\{j \mid (j+0.5)\delta \in A, \ j = 0,1\ldots,T\delta^{-1} - 1\}, A \in \mathcal{B}([0,1])$.

$$f_1^\delta(t) = \begin{cases} 0 & \text{if } t \neq (j+0.5) \\ \delta^{-1}\int_{j\delta}^{(j+1)\delta} f_1(t)\mathrm{d}x & \text{if } t = (j+0.5)\delta. \end{cases} \tag{8}$$

We define the time reversed counting processes as

$$N_i^\delta(t) = I\left(T - X_i^\delta \leq t\right), \quad (i = 1,\ldots,n),$$

with respect to the filtration

$$\mathcal{F}_t^{i,\delta} = \sigma\left(\left\{(T - X_i^\delta) \leq s : \ s \leq t\right\} \cup \mathcal{N}\right).$$

Similar to the continuous case one derives that the intensity of the counting process is

$$\nu_i^\delta(t) = \alpha^\delta(t) Z_i^\delta(t),$$

where the hazard ratio α, and the predictable filtering process, Z_i^δ, are

$$\alpha^\delta(t) = \frac{f_1^\delta(T-t)}{F_1^\delta(T-t)} = \frac{f_1^{\delta,R}(t)}{S_1^{\delta,R}(t)}, \quad Z_i^\delta(t) = I\{Y_i^\delta \le t \le (T - X_i^\delta)\}.$$

This means that also the discrete observation can be translated into Aalen's multiplicative intensity model. The main difference to the continuous case is that the Lebesgue measure is replaced by the counting measure, μ, which lives on a grid according to the aggregation level of the data. For the development of the theoretical properties of the discrete estimators in the next section we introduce the following functions for $t \in [0.5\delta, T - 0.5\delta]$,

$$\overline{\alpha}^\delta(t) = \frac{\delta^{-1} \int_{t-0.5\delta}^{t+0.5\delta} f(s)\mathrm{d}s}{\int_{t-0.5\delta}^{1} f(s)\mathrm{d}s}, \quad \overline{Z}_i^\delta(t) = I\{Y_i - 0.5\delta \le t \le (T - X_i) + 0.5\delta\}.$$

Note that $\overline{\alpha}^\delta(x_j) = \alpha^\delta(x_j)$ and $\overline{Z}_i^\delta(x_j) = Z_i^\delta(x_j)$. For δ converging to zero, we have that

$$\sup_{s\in[0,T]} |\overline{Z}^\delta(s)/n - \gamma(s)| = o_p(1). \tag{9}$$

3.2. The histogram estimator and chain ladders development factors

Let us assume that one chooses a bandwidth $h = c\delta$, $c = 1, 2, \ldots$, as bin width. Then, for t in the bin $[c_1, c_2)$, with width h, the histogram estimator of the previous section translates to

$$\widehat{\alpha}_h^{H,\delta}(t) = \frac{h^{-1} \sum_{i=1}^{n} \int_{c_1}^{c_2} \mathrm{d}N_i^\delta(s)}{\sum_{i=1}^{n} \int_{c_1}^{c_2} Z_i^\delta(s)\mathrm{d}\mu(s)} = \frac{O^{H,\delta}(t)}{E^{H,\delta}(t)}. \tag{10}$$

We will suppress the subscript h in $\widehat{\alpha}_h^{H,\delta}$, if $h = \delta$. Note that while the nominator equals the one from the continuous estimator in (2), the denominators are different. The reason is that when considering discrete observations, the exposure in the considered bins is not observed and hence needs to be estimated, see also Hoem (1969).

Proposition 3.2 Assume that assumption (S) and (D) hold, then for $h = \delta$, $\widehat{\alpha}_h^{H,\delta}(x_j)$ is an unbiased estimator of $\alpha^\delta(x_j)$. The estimation error is asymptotically normal with variance $(nh)^{-1}\gamma^{-1}(t_j)\alpha(t_j)$.
For $h = c\delta$, $c = 2, 3, \ldots$, it holds that

$$(nh)^{1/2}\{\widehat{\alpha}_h^H(x_j) - \alpha(x_j) - B(x_j)\} \xrightarrow{\mathcal{D}} N\{0, \sigma^2(x_j)\},$$

where

$$B(t) = \frac{1}{24} f^{R''}(x_j)\{S^R(t_j)\}^{-1}\delta^2 + (\overline{\alpha}^\delta)'(x_j)\left\{(x_j - c_1) - \frac{1}{2}h\right\} + o(h + \delta^2),$$

$$\sigma^2(t) = \overline{\alpha}^\delta(x_j)\gamma(x_j)^{-1}.$$

The proof can be found in the Appendix C.

We now discuss the relationship between chain ladder's development factors and the discrete histogram estimator of the hazard, $\widehat{\alpha}_h^{H,\delta}$, when one chooses that the bin-width h equals the discretization δ. Note that the development factor, (7), can be rewritten as

$$\widehat{\lambda}_j = \frac{E^{H,\delta}(x_j)}{E^{H,\delta}(x_j) - \delta O^{H,\delta}(x_j)}. \tag{11}$$

Theorem 3.3 Assume that λ_j is the j-th development factor derived from the chain ladder algorithm. It holds that

$$\widehat{\lambda}_j = \frac{1}{1 - \delta \; \widehat{\alpha}^{H,\delta}(T - x_j)}. \tag{12}$$

Furthermore, it holds that

$$\widehat{\lambda}(x_j) = 1 + \delta \; \widehat{\alpha}^{H,\delta}(T - x_j) + O_p(\delta^2). \tag{13}$$

Proof This follows directly from (10) and (11).

Equation (12) tells us that there is an exact and deterministic relationship between the histogram estimator and the development factor. Equation (13) even gives asymptotic equality when the development factors are subtracted by 1.

We conclude the following. In continues time, chain ladders development factors and a histogram estimator of the hazard in reversed time are the same entity. Or in other words the development factors aim to estimate a hazard in reversed time via a histogram approach. To make this clear, we introduce the new notation

$$\widehat{\lambda}^{H,\delta}(x_j) = \widehat{\lambda}_j = \frac{1}{1 - \delta \; \widehat{\alpha}^{H,\delta}(T - x_j)}. \tag{14}$$

When working in the continuous setting, or say daily level, those classical development factors will be too noisy. Thus, one will need to increase, h or equivalently δ, to increase performance. A better alternative might be to replace the classical development-factors by kernel smoothed versions.

In the next section we introduce discrete versions of the local constant and local linear kernel estimator.

3.3. Local polynomial estimator

It is straight forward to see that the solutions of the discrete versions of (3) and (5) are given by

$$\widehat{\alpha}_{h,K}^{LC,\delta}(t) = \frac{h^{-1}\sum_{i=1}^{n}\int K_h(t-s)\mathrm{d}N_i^{\delta}(s)}{\sum_{i=1}^{n}\int K_h(t-s)Z_i^{\delta}(s)\mathrm{d}\mu(s)} = \frac{O^{LC,\delta}(t)}{E^{LC,\delta}(t)}. \tag{15}$$

and

$$\widehat{\alpha}_{h,K}^{LL,\delta}(t) = \frac{\sum_{i=1}^{n}\int K_h(t-s)\left\{a_2^{\delta}(t) - a_1^{\delta}(t)(t-s)\right\}\,\mathrm{d}N_i^{\delta}(s)}{\sum_{i=1}^{n}\int K_h(t-s)\left\{a_2^{\delta}(t) - a_1^{\delta}(t)(t-s)\right\}Z_i^{\delta}(s)\,\mathrm{d}\mu(s)} = \frac{O^{LL,\delta}(t)}{E^{LL,\delta}(t)}, \tag{16}$$

where

$$a_j^{\delta}(t) = n^{-1}\int K_h(t-s)(t-s)^j Z^{\delta}(s)d\mu(s) \quad (j=0,1,2).$$

Proposition 3.4 Assume that assumption (S) and (D) are satisfied. Then, the following asymptotics holds for $t \in (0,T)$:

$$(nh)^{1/2}\left\{\widehat{\alpha}_{h,K}^{LC}(t) - \alpha(t) - B_{LC,\delta}(t)\right\} \xrightarrow{\mathfrak{D}} N\left\{0, \sigma_{LC,\delta}^2(t)\right\},$$

$$(nh)^{1/2}\left\{\widehat{\alpha}_{h,K}^{LL}(t) - \alpha(t) - B_{LL,\delta}(t)\right\} \xrightarrow{\mathfrak{D}} N\left\{0, \sigma_{LL,\delta}^2(t)\right\},$$

where

$$B_{LC,\delta}(x) = \frac{1}{24}f^{R''}(x_j)\{S^R(x_j)\}^{-1}\delta^2 + \mu_2(K)h^2\left\{(\overline{\alpha}^{\delta})'(x_j)\gamma'(x_j)\gamma^{-1}(x_j) + \frac{1}{2}(\overline{\alpha}^{\delta})''(x_j)\right\} + o(\delta^2 + h^2),$$

$$B_{LL,\delta}(x) = \frac{1}{24}f^{R''}(x_j)\{S^R(x_j)\}^{-1}\delta^2 + \frac{1}{2}\mu_2(K)h^2(\overline{\alpha}^{\delta})''(x_j) + o(\delta^2 + h^2),$$

$$\sigma_{LC,\delta}^2(x) = \sigma_{LL,\delta}^2(x) = R(K)\alpha^{\delta}(t)\gamma(x)^{-1}.$$

The proof can be found in the Appendix C.

We now define the local constant and local linear development factors which can be used in the chain ladder approach.

$$\widehat{\lambda}^{LC,\delta}(x_j) = \frac{1}{1 - \delta\,\widehat{\alpha}_h^{LC,\delta}(T - x_j)}, \quad \widehat{\lambda}^{LL,\delta}(x_j) = \frac{1}{1 - \delta\,\widehat{\alpha}_h^{LL,\delta}(T - x_j)}. \tag{17}$$

4. Simulation study

To illustrate the finite sample performance, we simulated three models assuming independent underwriting and delay components. For simplicity we set $T = 1$. For the development component, in model 1 and 2, we chose that $X \sim Beta(2,5)$, and in the third model we chose a more steep development pattern with $X \sim Exponential(5)$. For the underwriting variable, Y, in the first model, we assume a uniform distribution and

hence the chain ladder assumptions are satisfied for every aggregation. In the second and third model the density of Y is linearly increasing, i.e., $f_2(y) = 2y$. This means that the aggregated approaches will estimate a biased reserve. We also tried many other distributions but they did not change the conclusions we derived from those three models presented here.

We have run 500 repetitions with sample-sizes of $n = 200, 1000, 5000, 10000$ to estimate the relative error, $error = (E[R] - \widehat{R})/E[R]$, where $\widehat{R}$ is the reserve estimate derived by the chain ladder algorithm using the development factors in (14) and (17).

This is done by calculating the chain ladder development factors for aggregation levels $\delta \in \{0.01, 0.02, 0.04, 0.1, 0.2\}$. For $\delta = 0.01$ we also calculated the local linear and the local constant versions. The discretization $\delta = 0.01$ should approximate the continuous model well enough for the smaller sample sizes (n=200, 1000). For the greater sample sizes (n=5000, 10000), the performance of the local polynomial estimators could have been improved with a smaller δ, according to the asymptotic theory in the previous section. The way the code is implemented, computation time depends on the aggregation level, δ, and is unaffected by the sample size, n. More details are given in the Appendix.

This paper does not discuss the problem of how to choose a bandwidth b. This issue needs to be addressed separately where a cross-validation procedure is developed and assessed. Depending on the estimation purpose, i.e. if one interested in the full sum of the lower triangle or in the diagonal sums of the lower triangle, one will have different loss functions with different optimal bandwidths. In this simulation study, we have used the bandwidth optimal for the given loss function in each simulation step. This choice is infeasible in practice. To give an idea about the robustness of the estimators with respect to the bandwidth, for the local constant estimator, we also included a bandwidth which is randomly picked in every simulation step from a quite wide range depending on model and sample size. An eye picked or cross-validated bandwidth is then expected to have a performance in between the optimal and random choice. For the local polynomial estimators we have used the Epanechnikov kernel, as kernel K.

In Table 1 and Figure 3, we see that the histogram estimator becomes better the more one aggregates. This is consistent with the theory, since there is no bias in the estimation and accuracy can hence be reduced by aggregation via reduction of variance. But even when aggregated to a triangle with only 5 periods ($\delta = 0.2$) the kernel estimators are competitive, and the local constant estimator is even favorable. A change in sample size does not seem to alter the conclusion but improves the estimators uniformly.

The results of the second simulation study are presented in Table 2. Here, one can see that the choice of the aggregation for the chain ladder approach is a classical bias variance trade off. The results are also visualized in via boxplots in Figure 4. Also in this model the conclusion is to prefer the kernel estimators.

Similar results are given in the third model, Table 3 and Figure 5, which indicates that the results are independent of the distribution choice and also hold in a harder estimation problem with a sharp decay of mass.

An interesting result is that independent of the models, extreme estimation errors are always overestimating the reserve. This is independent of the true distributions or the way the development factors are estimated but seems to be a feature of the chain ladder

16

technique.

In both models the local linear estimator performed surprisingly bad compared to the local constant estimator, but might be better in other scenarios.

n		$\widehat{\lambda}^{LL,0.01}_{b.opt}$	$\widehat{\lambda}^{LC,0.01}_{b.opt}$	$\widehat{\lambda}^{LC,0.01}_{b.random}$	$\widehat{\lambda}^{H,0.01}$	$\widehat{\lambda}^{H,0.02}$	$\widehat{\alpha}^{H,0.04}$	$\widehat{\lambda}^{H,0.1}$	$\widehat{\lambda}^{H,0.2}$
200	Mean	-1.746	-0.590	0.217	-1.311	-1.146	-1.217	-0.765	-0.474
	Median	-0.089	-0.053	0.607	1.042	0.999	0.848	0.082	-0.178
	SD	3.674	1.289	2.825	9.033	8.177	7.498	4.528	2.957
1000	Mean	-0.906	-0.105	0.536	-0.234	-0.304	-0.313	-0.146	-0.088
	Median	-0.082	-0.010	0.601	0.601	0.479	0.243	0.059	-0.029
	SD	1.850	0.364	1.218	3.845	3.390	2.625	1.631	1.261
5000	Mean	-0.514	-0.004	0.801	-0.115	-0.076	-0.090	-0.022	-0.004
	Median	-0.059	-0.001	0.788	0.348	0.246	0.046	0.015	-0.001
	SD	1.090	0.083	0.559	1.785	1.335	1.088	0.709	0.544
10000	Mean	-0.429	0.005	0.804	-0.045	-0.012	-0.032	-0.007	0.000
	Median	-0.084	0.003	0.814	0.214	0.102	0.027	0.015	-0.014
	SD	0.813	0.079	0.446	1.163	0.932	0.748	0.511	0.388

Table 1: Simulation results over 500 repetitions for the relative estimation error of the reserve. The development delay, X, has a Beta distribution with parameters $(2, 5)$, and the underwriting date density, Y, is uniformly distributed. Sample size is $n = 200, 1000, 5000, 10000$. For $n = 200, 1000$: b.random $\in [0.05, 0.3]$, for $n = 1000, 5000$: $b.random \in [0.05, 0.25]$

5. Concluding remarks

In this paper we connected classical chain ladder to the continues chain ladder model of Martínez-Miranda et al. (2013). We derive a one to one connection between the development factors and a histogram hazard estimator and then improve this histogram estimator by more efficient kernel smoothers. However, the hazard interpretation also allows for straight forward generalisations to more flexible models allowing for calendar time effects and covariates with specific claim informations. This can be done by extending the univariate hazard estimation case to the multivariate case.

Another point is that we only considered claim counts. A generalisation suitable for claim amounts is explained in Hiabu (2016). The assumption necessary hereby is that the influences of development delay and underwriting date on the claim severity are independent to each other. If this holds, then everything done in this paper can be done in the same way with claim amounts. The uncertainty will, however, depend on the severity distribution. With this generalisation it would be interesting to compare the methods in this paper with other individual reserving models like Martínez-Miranda, Nielsen, and Verrall (2012) or Antonio and Plat (2014).

In this the paper, we have derived asymptotic results for the estimation uncertainty of the hazard/development factors. Uncertainty of the reserve or estimated sum of the
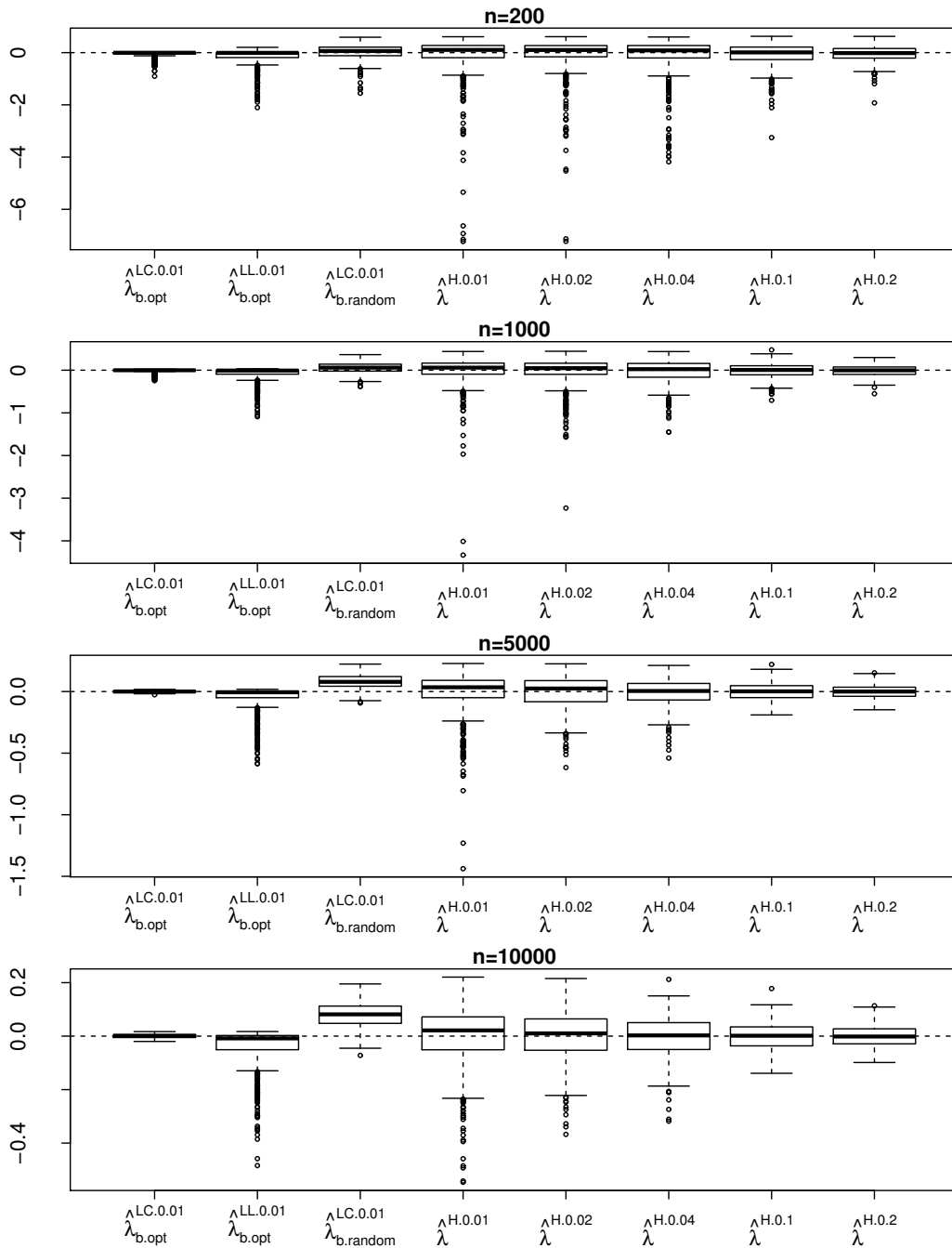
Figure 3: Boxplot results over 500 repetitions for the relative estimation error of the reserve. The development delay, X, has a Beta distribution with parameters $(2,5)$, and the underwriting date density, Y, is uniformly distributed. Sample size is $n = 200, 1000, 5000, 10000$. For $n = 200, 1000$: b.random $\in [0.05, 0.3]$, for $n = 1000, 5000$: $b.random \in [0.05, 0.25]$
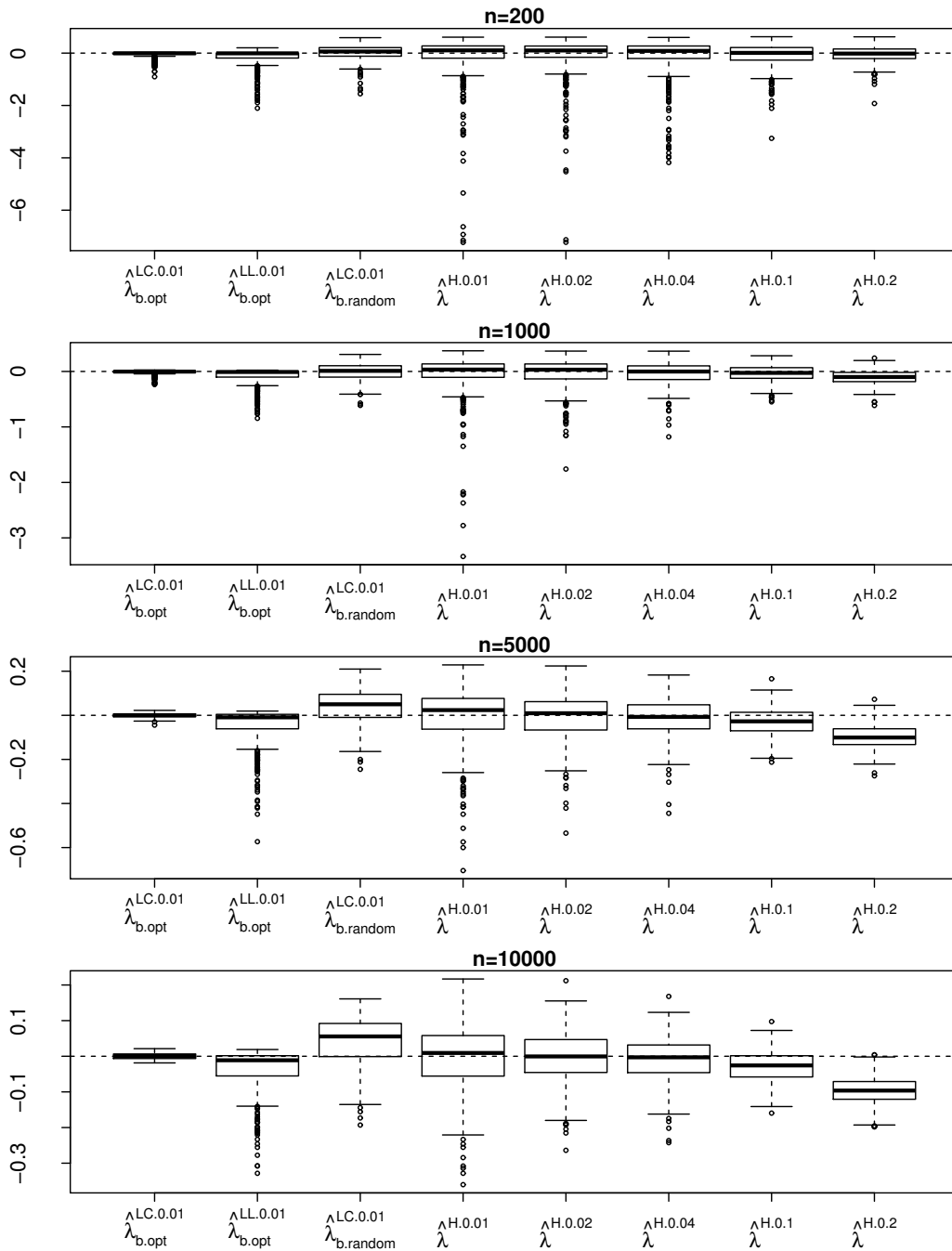
Figure 4: Boxplot results over 500 repetitions for the relative estimation error of the reserve. The development delay, X, has a Beta distribution with parameters $(2, 5)$, and the underwriting date density, Y, is linear increasing, $f_2(y) = 2y$. For $n = 200, 1000$: b.random $\in [0.05, 0.3]$, for $n = 1000, 5000$: $b.random \in [0.05, 0.25]$
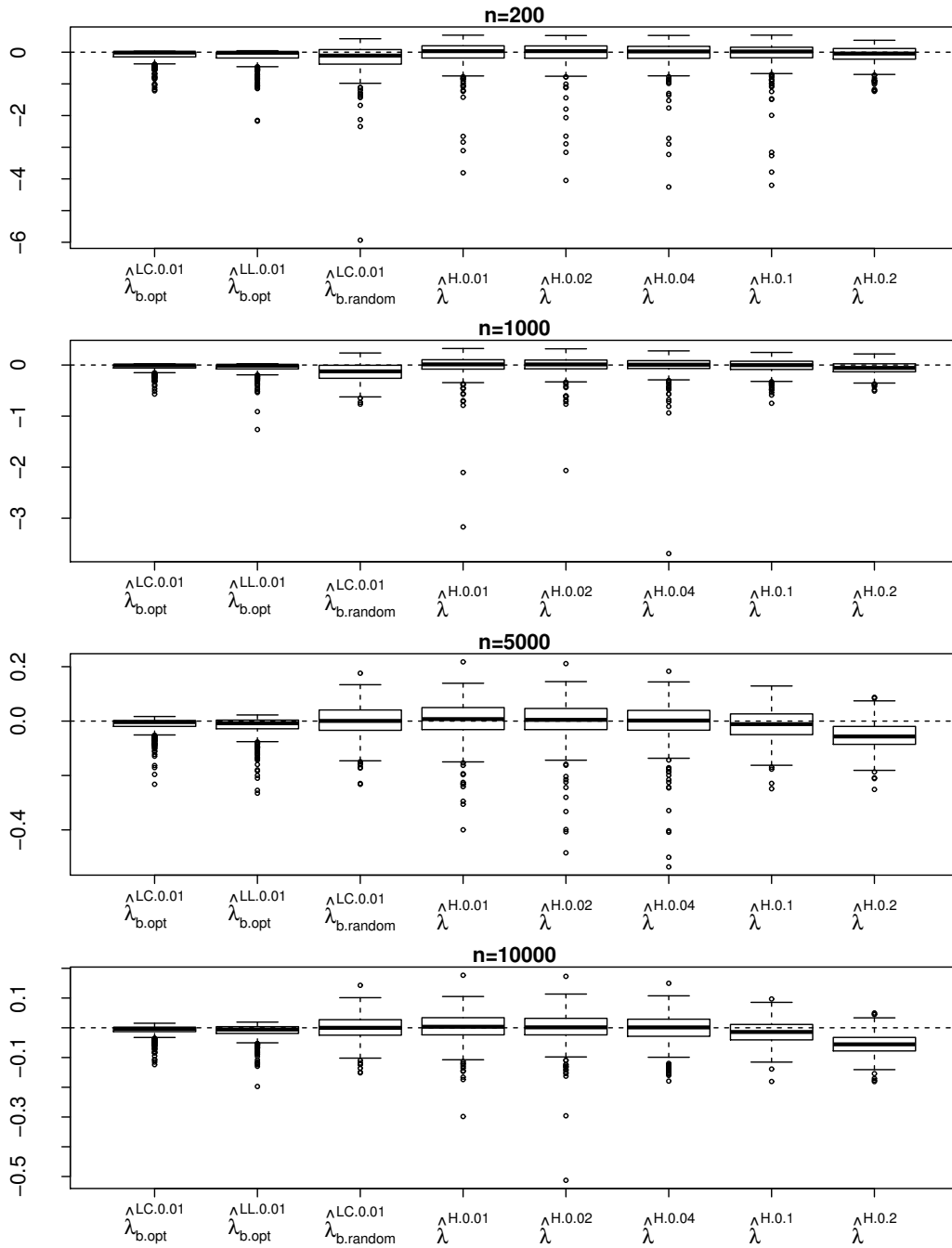
Figure 5: Boxplot results over 500 repetitions for the relative estimation error of the reserve. The development delay, X, has a exponential distribution with rate 5, and the underwriting date density Y, is linear increasing, $f_2(y) = 2y$. For $n = 200, 1000$: b.random $\in [0.05, 0.25]$, for $n = 1000, 5000$: $b.random \in [0.01, 0.1]$

n		$\widehat{\lambda}^{LL,0.01}_{b.opt}$	$\widehat{\lambda}^{LC,0.01}_{b.opt}$	$\widehat{\lambda}^{LC,0.01}_{b.random}$	$\widehat{\lambda}^{H,0.01}$	$\widehat{\lambda}^{H,0.02}$	$\widehat{\alpha}^{H,0.04}$	$\widehat{\lambda}^{H,0.1}$	$\widehat{\lambda}^{H,0.2}$
200	Mean	-1.746	-0.590	0.217	-1.311	-1.146	-1.217	-0.765	-0.474
	Median	-0.089	-0.053	0.607	1.042	0.999	0.848	0.082	-0.178
	SD	3.674	1.289	2.825	9.033	8.177	7.498	4.528	2.957
1000	Mean	-0.824	-0.125	-0.062	-0.401	-0.374	-0.373	-0.352	-1.043
	Median	-0.094	-0.030	0.106	0.343	0.321	-0.017	-0.258	-0.999
	SD	1.546	0.365	1.500	3.604	2.622	2.051	1.414	1.271
5000	Mean	-0.464	-0.000	0.400	-0.084	-0.059	-0.120	-0.286	-0.975
	Median	-0.092	-0.001	0.500	0.237	0.094	-0.072	-0.273	-1.008
	SD	0.867	0.093	0.743	1.264	1.020	0.840	0.602	0.534
10000	Mean	-0.351	0.000	0.410	-0.053	-0.033	-0.092	-0.275	-0.958
	Median	-0.112	0.005	0.556	0.092	-0.005	-0.029	-0.258	-0.963
	SD	0.584	0.087	0.665	0.846	0.705	0.584	0.423	0.369

Table 2: Simulation results over 500 repetitions for the relative estimation error of the reserve. The development delay, X, has a Beta distribution with parameters $(2, 5)$, and the underwriting date density, Y, is linear increasing, $f_2(y) = 2y$. For $n = 200, 1000$: b.random $\in [0.05, 0.3]$, for $n = 1000, 5000$: $b.random \in [0.05, 0.25]$

lower triangle is not discussed in this paper. An analytic derivation seems not to be straightforward, since even if the true development factors are known, chain ladder uses the observed values to project into the lower triangle. However, since we are in a full statistical model, one could develop and implement a bootstrap approach which can also include parameter uncertainty. This would also be possible in the more general framework of Hiabu (2016) which is suitable for claim amounts.

A. Computational complexity

In this section we give a brief and not so scientific outline about the computational cost involved in the chain ladder algorithm implemented for this paper. The complexity does hereby not depend on the sample size but only on the dimension of the triangle, i.e, the number $(T\delta^{-1})$. In Table 4 we provide an idea of the computational complexity of the algorithm running in a standard computer (Intel(R) Core(TM) i5-4590S with 3.00 GHz and 8.00 GB-RAM with R working under Windows 7-64 bit). Specifically we have evaluated the run-time of one arbitrary simulated sample with $(T\delta^{-1}) = 100, 1000, 10000$. We have hereby split the computation time in the three different categories. Firstly, the aggregation from a triangle of size $T\delta^{-1}$ to a smaller triangle (Aggregation); note that it does hereby not matter to which size the triangle is aggregated. Secondly the calculation of the development factors via the different methods $(\widehat{\lambda}^H, \widehat{\lambda}^{LC}, \widehat{\lambda}^{LL})$ and lastly the chain ladder algorithm when the development factors are given (CL algorithm).

n		$\widehat{\lambda}^{LL,0.01}_{b.opt}$	$\widehat{\lambda}^{LC,0.01}_{b.opt}$	$\widehat{\lambda}^{LC,0.01}_{b.random}$	$\widehat{\lambda}^{H,0.01}$	$\widehat{\lambda}^{H,0.02}$	$\widehat{\alpha}^{H,0.04}$	$\widehat{\lambda}^{H,0.1}$	$\widehat{\lambda}^{H,0.2}$
200	Mean	-1.348	-1.033	-1.772	-0.437	-0.454	-0.539	-0.609	-0.759
	Median	-0.170	-0.126	-1.061	0.326	0.334	0.238	0.218	-0.434
	SD	2.518	1.946	4.328	4.101	4.208	4.194	4.353	2.744
1000	Mean	-0.553	-0.426	-1.400	-0.111	-0.055	-0.148	-0.200	-0.620
	Median	-0.120	-0.073	-1.242	0.125	0.113	0.062	-0.017	-0.567
	SD	1.117	0.800	1.790	2.242	1.731	2.209	1.364	1.198
5000	Mean	-0.188	-0.143	0.005	0.032	0.021	-0.032	-0.133	-0.539
	Median	-0.080	-0.037	0.006	0.073	0.046	0.019	-0.113	-0.565
	SD	0.375	0.298	0.582	0.670	0.708	0.740	0.554	0.506
10000	Mean	-0.123	-0.094	-0.001	0.022	0.008	-0.023	-0.142	-0.543
	Median	-0.058	-0.031	-0.001	0.034	0.015	0.012	-0.140	-0.556
	SD	0.255	0.199	0.422	0.475	0.522	0.468	0.382	0.352

Table 3: Simulation results over 500 repetitions for the relative estimation error of the reserve. The development delay, X, has a exponential distribution with rate 5, and the underwriting date density Y, is linear increasing, $f_2(y) = 2y$. For $n = 200, 1000$: b.random $\in [0.05, 0.25]$, for $n = 1000, 5000$: $b.random \in [0.01, 0.1]$

B. A martingale CLT

In Ramlau-Hansen (1983), the author presented a central limit theorem for the martingale $M(t) = N(t) - \int \alpha(t)Z(t)\mathrm{d}t$. This is essential to derive asymptotic normality of the kernel and also the histogram estimator of the hazard function α. As mentioned in that paper this central limit theorem is only a special case of Corollary 2 in Liptser and Shiryayev (1981) which also covers the discrete setting of chapter 3 in this paper. The result can be stated as follows.

Theorem B.1 For the continuous case: Consider a predictable process $W_n(t)$ and assume that for some $\sigma^2 \geq 0$ the following conditions are satisfied:

$$\int W_n^2(t)Z(t)\alpha(t)\mathrm{d}t = \sigma^2 + o_p(1),$$

$$\int W_n^2(t)I\left\{W_n^2(t) > \varepsilon\right\} Z(t)\alpha(t)\mathrm{d}t = o_p(1) \quad \text{for all } \varepsilon > 0.$$

Then, it holds that $\int W_n(u)\mathrm{d}M(u) \to N(0, \sigma^2)$, in distribution. For the discrete case: Consider a predictable process $H_n(t)$ and assume that for some $(\sigma^\delta)^2 \geq 0$ the following conditions are satisfied:

$$\int H_n^2(t)Z^\delta(t)\alpha^\delta(t)\mathrm{d}\mu(t) = \sigma^2 + o_P(1)$$

$$\int H_n^2(t)I\{H^2(t) > \varepsilon\}Z^\delta(t)\alpha^\delta(t)d\mu(t) = o_P(1) \quad \text{for all } \varepsilon > 0,$$

$T\delta^{-1}$	Aggregation	CL algorithm	$\widehat{\lambda}^H$	$\widehat{\lambda}^{LC}$	$\widehat{\lambda}^{LL}$
100	0.011	0.036	0.001	0.106	0.1109
1000	1.223	3.382	1.032	8.976	6.274
10000	117	1351	4659	5176	4933
	(2min)	(23min)	(78min)	(86min)	(82min)
Complexity	$(T\delta^{-1})^2$	$(T\delta^{-1})^3$	$(T\delta^{-1})^3$	$(T\delta^{-1})^3\mathcal{B}$	$(T\delta^{-1})^3\mathcal{B}$

Table 4: Computation time in seconds and complexity for the aggregation, the chain ladder algorithm and the development factor estimators. The local polynomial estimators also depend on the number of bandwidths $\mathcal{B}$. The running time for the LC and LL estimators are given for a choice with $\mathcal{B} = 50$.

then it holds that

$$\int H_n(t)\mathrm{d}M^\delta(t) \to N(0,(\sigma^\delta)^2),$$

where $M^\delta(t) = N^\delta(t) - \int \alpha^\delta(t)Z^\delta(t)\mathrm{d}\mu(t)$.

C. Proofs

C.1. Proof of Proposition 2.1

By defining

$$\alpha^*(t) = \frac{\sum_{i=1}^n \int_{c_1}^{c_2} \mathrm{d}\Lambda_i(s)}{\sum_{i=1}^n \int_{c_1}^{c_2} Z_i(s)\mathrm{d}s},$$

we divide the estimation error $\widehat{\alpha}_h^H(t) - \alpha(t)$ into a deterministic part, $\alpha^*(t) - \alpha(t)$, and a variable part, $\widehat{\alpha}_h^H(t) - \alpha^*(t)$. By a first order Taylor expansion we get for the deterministic part that

$$\alpha^*(t) - \alpha(t) = \frac{\sum_{i=1}^n \int_{c_1}^{c_2} \{\alpha(s) - \alpha(t)\}Z_i(s)\mathrm{d}s}{\sum_{i=1}^n \int_{c_1}^{c_2} Z_i(s)\mathrm{d}s} = \alpha'(t)h^{-1}\int_{c_1}^{c_2}(t-s)\ \mathrm{d}s + o(h).$$

For the variable part we have

$$\widehat{\alpha}_h^H(t) - \alpha^*(t) = \frac{\sum_{i=1}^n \int_{c_1}^{c_2} \mathrm{d}M_i(s)}{\sum_{i=1}^n \int_{c_1}^{c_2} Z_i(s)\mathrm{d}s}.$$

From (1) it directly follows that that the second condition of Theorem B.1 in the Appendix is satisfied for $W(s) = (nh)^{1/2}I(s \in [c1,c2])\{\int_{c_1}^{c_2} Z(s)\mathrm{d}s\}^{-1}$. To calculate the asymptotic variance several first order Taylor expansions of $\gamma(s)$ and $\alpha(s)$ yield

$$\int W^2(s)\alpha(s)Z(s)\mathrm{d}s = \alpha(t)\gamma(t)^{-1} + o(1),$$

23

where here and below, the integral, $\int$, with no limits denotes integration over the whole support, that is $\int_0^T$. We deduce that $\widehat{\alpha}_h^H(t) - \alpha^*(t)$ is centered and asymptotically normal with variance $\sigma^2(t)$.

C.2. Proof of Proposition 3.2

By defining

$$\alpha^*(t_j) = \frac{\sum_{i=1}^n \int_{c_1}^{c_2} \mathrm{d}\Lambda_i^\delta(s)}{\sum_{i=1}^n \int_{c_1}^{c_2} Z_i^\delta(s)\mathrm{d}\mu(s)},$$

we divide the estimation error $\widehat{\alpha}_h^H(t_j) - \alpha(t_j)$ into a deterministic part, $\alpha^*(t_j) - \alpha(t_j)$, and a variable part, $\widehat{\alpha}_h^H(t_j) - \alpha^*(t_j)$. By a first order Taylor expansion we get for the deterministic part that

$$\alpha^*(t_j) - \alpha(t_j) = \frac{\sum_{i=1}^n \int_{c_1}^{c_2} \{\alpha^\delta(s) - \alpha(t_j)\} Z_i^\delta(s)\mathrm{d}\mu(s)}{\sum_{i=1}^n \int_{c_1}^{c_2} Z_i^\delta(s)\mathrm{d}\mu(s)}$$

$$= \alpha^\delta(t_j) - \alpha(t_j) + (\overline{\alpha}^\delta)'(t_j)h^{-1}\int_{c_1}^{c_2}(t_j - s)\,\mathrm{d}s + o(h)$$

$$= \frac{1}{24}f^{R''}(t_j)\{S^R(t_j)\}^{-1}\delta^2 + (\overline{\alpha}^\delta)'(t_j)\left\{(t_j - c_1) - \frac{1}{2}h\right\} + o(h + \delta^2).$$

For the variable part we have

$$\widehat{\alpha}_h^{H,\delta}(t_j) - \alpha^*(t_j) = \frac{\sum_{i=1}^n \int_{c_1}^{c_2} \mathrm{d}M_i^\delta(s)}{\sum_{i=1}^n \int_{c_1}^{c_2} Z_i^\delta(s)\mathrm{d}s}.$$

From (9) it directly follows that that the second condition of Theorem B.1 in the Appendix is satisfied for $H(s) = (nh)^{1/2}I(s \in [c1, c2])\{\int_{c_1}^{c_2} Z^\delta(s)\mathrm{d}\mu(s)\}^{-1}$. To calculate the asymptotic variance several first order Taylor expansions of $\gamma(s)$ and $\overline{\alpha}^\delta(s)$ yield

$$\int H^2(s)\overline{\alpha}^\delta(s)Z^\delta(s)\mathrm{d}\mu(s) = \overline{\alpha}^\delta(t_j)\gamma(t_j)^{-1} + o(1).$$

We deduce that $\widehat{\alpha}_h^{H,\delta}(t) - \alpha^*(t)$ is centered and asymptotically normal with variance $\sigma^2(t)$.

C.3. Proof of Proposition 3.4

We only show the result for local constant estimator. The case for the local linear estimator is proved in the same way after the kernel $\overline{K}_{t,h}(t - s)$ is replaced by $K(t - s)Z^{-1}(s)$. In Nielsen and Tanggaard (2001) it was shown that this can be done when studying pointwise first order asymptotics. We define

$$\alpha^*(t_j) = \frac{\sum_{i=1}^n \int K_h(t_j - s)\mathrm{d}\Lambda_i^\delta(s)}{\sum_{i=1}^n \int K_h(t_j - s)Z_i^\delta(s)\mathrm{d}\mu(s)}.$$

24

The estimation error can then be divided into into a deterministic part, $\widehat{\alpha}_h^H(t_j) - \alpha(t_j)$ $\alpha^*(t_j) - \alpha(t_j)$, and a variable part, $\widehat{\alpha}_h^H(t_j) - \alpha^*(t_j)$. By a second order Taylor expansion we get for the deterministic part that

$$
\begin{aligned}
\alpha^*(t_j) - \alpha(t_j) &= \frac{\int K_h(t_j - s)\{\alpha^\delta(s) - \alpha(t_j)\}Z_i^\delta(s)\mathrm{d}\mu(s)}{\int K_h(t_j - s)Z_i^\delta(s)\mathrm{d}\mu(s)} \\
&= \alpha^\delta(t_j) - \alpha(t_j) + (\overline{\alpha}^\delta)'(t_j)\frac{\int K_h(t_j - s)(t - s)Z(s)\mathrm{d}\mu(s)}{\int K_h(t_j - s)Z(s)\mathrm{d}\mu(S)} \\
&\quad + \frac{1}{2}(\overline{\alpha}^\delta)''(t_j)\frac{\int K_h(t_j - s)(t - s)^2 Z(s)\mathrm{d}\mu(s)}{\int K_h(t_j - s)Z(s)\mathrm{d}\mu(S)} + o(h^2) \\
&= \frac{1}{24}f^{R''}(t_j)\{S^R(t_j)\}^{-1}\delta^2 \\
&\quad + \mu_2(K)h^2\left\{(\overline{\alpha}^\delta)'(t_j)\gamma'(t_j)\gamma^{-1}(t_j) + \frac{1}{2}(\overline{\alpha}^\delta)''(t_j)\right\} + o(\delta^2 + h^2)
\end{aligned}
$$

For the variable part we have

$$
\widehat{\alpha}_h^{H,\delta}(t_j) - \alpha^*(t_j) = \frac{\sum_{i=1}^n \int K_h(t_j - s)\mathrm{d}M_i^\delta(s)}{\sum_{i=1}^n \int K_h(t_j - s)Z_i^\delta(s)\mathrm{d}s}.
$$

From (9) it directly follows that that the second condition of Theorem B.1 in the Appendix is satisfied for $H(s) = (nh)^{1/2}\{\int K_h(t_j - s)Z^\delta(s)\mathrm{d}\mu(s)\}^{-1}$. To calculate the asymptotic variance Taylor expansions of $\gamma(s)$ and $\overline{\alpha}^\delta(s)$ yield

$$
\int H^2(s)\overline{\alpha}^\delta(s)Z^\delta(s)\mathrm{d}\mu(s) = \overline{\alpha}^\delta(t_j)\gamma(t_j)^{-1} + o(1).
$$

We deduce that $\widehat{\alpha}_h^{H,\delta}(t) - \alpha^*(t)$ is centered and asymptotically normal with variance $\sigma^2(t)$.

C.4. Proof of Proposition 3.1

We have to show that Assumption D is a necessary and sufficient condition so that $P\left(X^\delta = x_j, Y^\delta = y_k\right)$ factorizes for $(x_j, y_k) \in \mathcal{I}^\delta$. We will only show that Assumption D is a necessary condition the sufficiency is easily shown by plugging in the solution in a similar manner. Consider the case for $x_j = 0.5\delta$. It holds that

$$
pr(X^\delta = 0.5\delta)pr(Y^\delta = y_k) = \sum_{l=0}^{T\delta^{-1}} \int_0^\delta f_1(x) \int_{l\delta}^{(l+1)\delta - x} f_2(y)\ \mathrm{d}y\mathrm{d}x \int_{k\delta}^{(k+1)\delta} f_2(y)\mathrm{d}y.
$$

We also have that

$$
pr\left(X^\delta = 0.5\delta, Y^\delta = y_k\right) = \int_0^\delta f_1(x) \int_{k\delta}^{(k+1)\delta - x} f_2(y)\mathrm{d}y\mathrm{d}x
$$

Without further restrictions on f_1 those two terms can only be equal if for almost every $x \in [0, \delta]$, and every k

$$\sum_{l=0}^{T\delta^{-1}} \int_{l\delta}^{(l+1)\delta-x} f_2(y) \; \mathrm{d}y \int_{k\delta}^{(k+1)\delta} f_2(y)\mathrm{d}y = \int_{k\delta}^{(k+1)\delta-x} f_2(y)\mathrm{d}y$$

We assume without loss of generality that there is a k where $\int_{k\delta-x}^{(k+1)\delta} f_2(y)\mathrm{d}y$ is not zero for all $x \in [0, \delta]$ (otherwise restrict the range of x). Fixing this k we conclude that

$$\sum_{l=0}^{T\delta^{-1}} \int_{l\delta}^{(l+1)\delta-x} f_2(y) \; \mathrm{d}y \Big/ \int_{k\delta}^{(k+1)\delta-x} f_2(y)\mathrm{d}y$$

does not depend on x, which shows the necessity.

References

Aalen, O. O. (1978). "Non-parametric inference for a family of counting processes". In: Ann. Stat. 6, pp. 701–726.

Addona, V., J. Atherton, and D. B. Wolfson (2012). "Testing the assumption of independence of truncation time and failure time". In: Int. J. Biostat. 8.

Andersen, P., O. Borgan, R. Gill, and N. Keiding (1993). Statistical Models Based on Counting Processes. New York: Springer.

Antonio, K. and R. Plat (2014). "Micro-level stochastic loss reserving for general insurance". In: Scand. Actuar. J 2014, pp. 649–669.

Arjas, E. (1989). "The claims reserving problem in non-life insurance: Some structural ideas". In: Astin Bulletin 19, pp. 139–152.

Borgan, Ø. (1984). "Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data". In: Scand. J. Stat, pp. 1–16.

Cleveland, W. S. (1979). "Robust locally weighted regression and smoothing scatterplots". In: J. Am. Stat. Assoc. 74, pp. 829–836.

Drieskens, D., M Henry, J.-F. Walhin, and J. Wielandts (2012). "Stochastic projection for large individual losses". In: Scand. Actuar. J. 2012, pp. 1–39.

England, P. D. and R. J. Verrall (2002). "Stochastic Claims Reserving In General Insurance". In: British Actuarial Journal 8, pp. 443–544.

Fan, J. and I. Gijbels (1996). Local polynomial modelling and its applications. London: Chapman and Hall.

Godecharle, E. and K. Antonio (2015). "Reserving by Conditioning on Markers of Individual Claims: A Case Study Using Historical Simulation". In: North American Actuarial Journal 19, pp. 273–288.

Hiabu, M. (2016). "Continuous chain-ladder with paid data: the theoretical foundation". In: Preprint.

Hiabu, M., E. Mammen, M. D. Martínez-Miranda, and J. P. Nielsen (2016). "In-sample forecasting with local linear survival densities". In: Preprint.

Hiabu, M., C. Margraf, M. D. Martínez-Miranda, and J. P. Nielsen (2015). "Cash flow generalisations of non-life insurance expert systems estimating outstanding liabilities". In: Expert Systems With Applications 45, pp. 400–409.

Hjort, N. L., M. West, and S. Leurgans (1992). "Semiparametric Estimation Of Parametric Hazard Rates". In: Survival Analysis: State of the Art. Ed. by J. P. Klein and P. Goel. Vol. 211. Nato Science, pp. 211–236.

Hoem, J. M. (1969). "Fertility rates and reproduction rates in a probabilistic setting". In: Biométrie-Praximétrie 10, pp. 38–66.

Huang, J., X. Wu, and X. Zhou (2016). "Asymptotic behaviors of stochastic reserving: Aggregate versus individual models". In: European J. Oper. Res. 249, pp. 657–666.

Jacod, J. (1979). Calcul stochastique et problemes de martingales. Berlin: Springer.

Jacod, J. and A. N. Shiryaev (1987). Limit Theorems for Stochastic Processes. Berlin: Springer.

Keiding, N. (1990). "Statistical inference in the Lexis diagram". In: PHILOS T ROY SOC A 332, pp. 487–509.

Kremer, E. (1982). "IBNR-claims and the two-way model of ANOVA". In: Scand. Actuar. J. 1982, pp. 47–55.

Kuang, D., B. Nielsen, and J. P. Nielsen (2009). "Chain-ladder as maximum likelihood revisited". In: Ann. Actuar. Sci 4, pp. 105–121.

Lagakos, S. W., L. M. Barraj, and V. De Gruttola (1988). "Nonparametric Analysis of Truncated Survival Data, with Application to AIDS". In: Biometrika 75, pp. 515–523.

Liptser, R. S. and A. N. Shiryayev (1981). "A functional central limit theorem for semimartingales". In: Theory Probab. Appl. 25, pp. 667–688.

Macaulay, F. R. (1931). The smoothing of time series. New York: NationalBureau of Economic Research.

Mack, T. (1993). "Distribution-free calculation of the standard error of chain ladder reserve estimates". In: Astin Bulletin 23, pp. 213–225.

Mandel, M. and R. A. Betensky (2007). "Testing goodness of fit of a uniform truncation model". In: Biometrics 63, pp. 405–412.

Martínez-Miranda, M. D., J. P Nielsen, S. Sperlich, and R. Verrall (2013). "Continuous Chain Ladder: Reformulating and generalising a classical insurance problem". In: Expert. Syst. Appl. 40, pp. 5588–5603.

Martínez-Miranda, M. D, J. P. Nielsen, and R. Verrall (2012). "Double Chain Ladder". In: Astin Bull. 42, pp. 59–76.

Nielsen, J. P. (1998). "Marker dependent kernel hazard estimation from local linear estimation". In: Scand. Actuar. J. 1998, pp. 113–124.

Nielsen, J. P. and C. Tanggaard (2001). "Boundary and bias correction in kernel hazard estimation". In: Scand. J. Stat. 28, pp. 675–698.

Norberg, R. (1993). "Prediction of Outstanding Liabilities in Non-Life Insurance". In: Astin Bull. 23, pp. 95–115.

Pigeon, M., K. Antonio, and M. Denuit (2013). "Individual loss reserving with the multivariate skew normal framework". In: Astin Bulletin 43, pp. 399–428.

Ramlau-Hansen, H. (1983). "Smoothing counting process intensities by means of kernel functions". In: Ann. Stat. 11, pp. 453–466.

Renshaw, A. E. and R. J. Verrall (1998). "A stochastic model underlying the chain-ladder technique". In: British Actuarial Journal 4, pp. 903–923.

Rosenlund, S. (2012). "Bootstrapping individual claim histories". In: Astin Bull. 42, p. 291.

Schiegl, M. (2015). "A model study about the applicability of the Chain Ladder method". In: Scand. Actuar. J. 2015, pp. 482–499.

Schmidt, K. D. (2012). "Loss prediction based on run-off triangles". In: AStA Adv. Stat. Anal. 96, pp. 265–310.

Silverman, B. W. (1986). Density estimation for statistics and data analysis. London: Chapman and Hall.

Simonoff, J. S. (1998). Smoothing methods in statistics. New York: Springer.

Stone, C. J. (1977). "Consistent nonparametric regression". In: Ann. Stat. 5, pp. 595–620.

Taylor, G. C. (1986). Claims reserving in non-life insurance. Amsterdam: Elsevier Science Ltd.

Tsai, W.–Y (1990). "Testing the assumption of independence of truncation time and failure time". In: Biometrika 77, pp. 169–177.

Verrall, R. J. (1991). "Chain ladder and maximum likelihood". In: J. Inst. Actuar. 118, pp. 489–499.

Verrall, R., J. P. Nielsen, and A. Jessen (2010). "Including Count Data in Claims Reserving". In: Astin Bull. 40, pp. 871–887.

Ware, J. H. and D. L. DeMets (1976). "Reanalysis of some baboon descent data". In: Biometrics 32, pp. 459–463.

Wüthrich, M. V. and M. Merz (2008). Stochastic claims reserving methods in insurance. West Sussex: John Wiley & Sons.

Wüthrich, M. V., M. Merz, and H. Bühlmann (2008). "Bounds on the estimation error in the chain ladder method". In: Scand. J. Stat. 2008, pp. 283–300.

Zhao, X. B. and X. Zhou (2010). "Applying copula models to individual claim loss reserving methods". In: Insurance Math. Econom 46, pp. 290–299.

Zhao, X. B., X. Zhou, and J. L. Wang (2009). "Semiparametric model for prediction of individual claim loss reserving". In: Insurance Math. Econom 45, pp. 1–8.