Cass Business School
CITY UNIVERSITY LONDON

Continuous Chain Ladder

Richard Verrall

Joint work with M.D. Martínez-Miranda, J.P. Nielsen and S. Sperlich

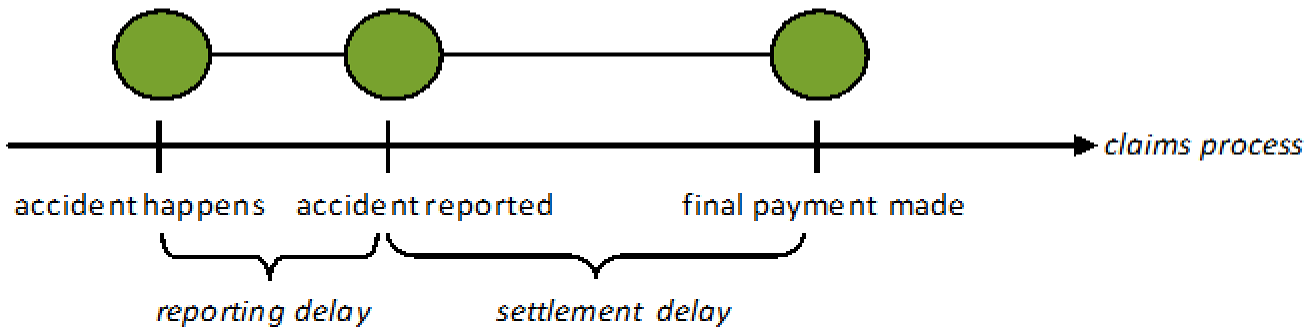
Cass Business School
CITY UNIVERSITY LONDON

The Claims Reserving exercise

- Claims are first notified and then (at a later date) settled - **reporting delays** and **settlement delays** exist.
- The amount and timing of future claims is not known and this creates an **uncertainty over the amount of reserves** that needs to be held.
- Companies have an **outstanding liability** for claims events that have already happened and for claims that have not yet been fully settled.

The life of an individual claim

Individual claims mechanism in the general claims process



Main components in the process:

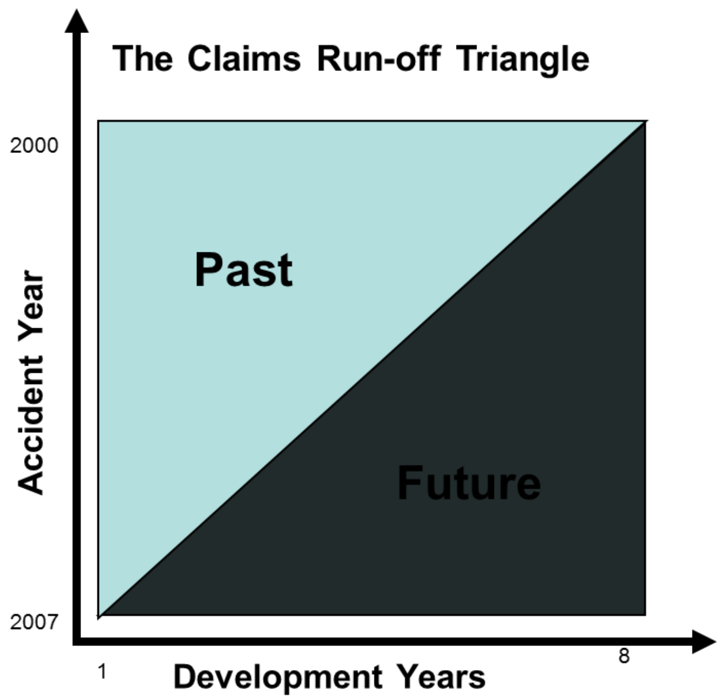
- **Reporting** delay
- **Settlement** delay (development process from reporting)
- Payments size

Type of claims: **IBNR**, **RBNS** and closed claims

Cass Business School
CITY UNIVERSITY LONDON

The data

- The available information matters: **look at the data…**
- Aggregated **run-off triangles** lead to classical collective methods such as the popular Chain Ladder method.



Accident (underwriting) year: year in which the claim arose or was underwritten

Development year: difference between the payment (or other action) year and the accident year

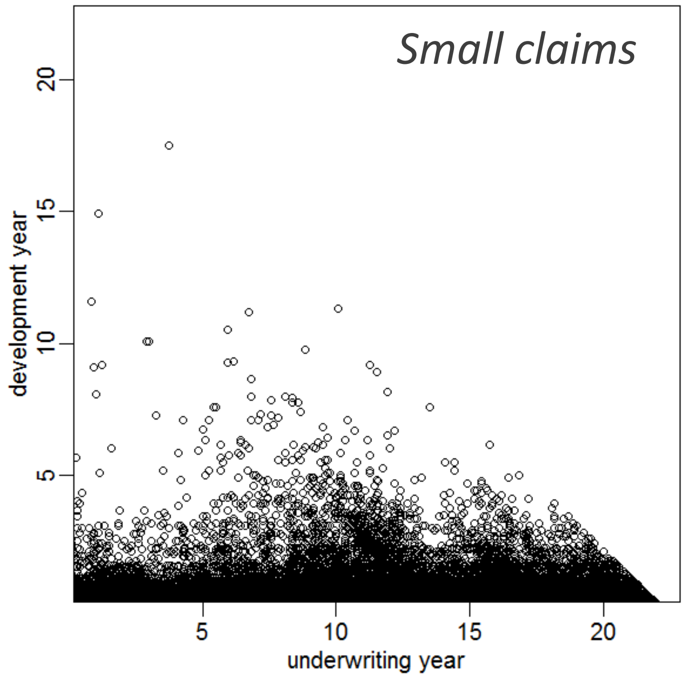
Periods: years, quarters …

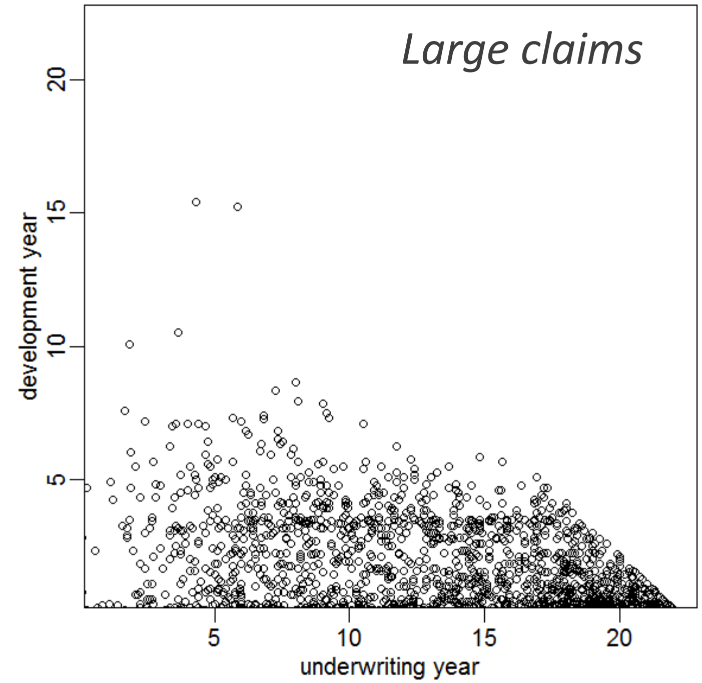
Data: payments, number of claims …

Cass Business School
CITY UNIVERSITY LONDON

When you have "more data": going granular

- **Micro-level data** leading to **individual claim loss models** (among others Taylor et al. 2008, Zhao and Zhou 2010, Antonio and Platz 2012)





Cass Business School
CITY UNIVERSITY LONDON

It is time to modernise claims reserving methodology

- Classical reserving methods rely on aggregate run-off triangles since only recently has micro-level information been available at companies.

- Now the the challenge is to use micro-level information in an efficient way.
- There is a growing awareness among non-life actuaries that modern statistical expert models should be used when analysing this type of data. However, there is no clear concensus on how to proceed.

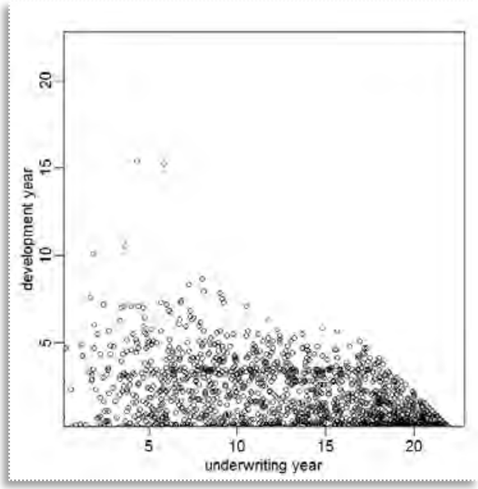
Cass Business School
CITY UNIVERSITY LONDON

Going granular in reserving…
…but respecting the chain ladder approach

- Maybe the best approach to developing new methodology is to **first understand what already exists.**
- We suggest **reformulating the classical chain ladder method** into a modern statistical framework. Then, a natural way to improve it will come: Continuous Chain Ladder.
- Some good reasons to proceed in such way:
  1. Actuaries have extensive **tacit knowledge.**
  2. When you build a system from many small systems you get **bias**. Keep the chain ladder mean as a benchmark.
  3. Simpler models are preferred for forecasting.

Cass Business School
CITY UNIVERSITY LONDON

Reformulating claims reserving as a density problem



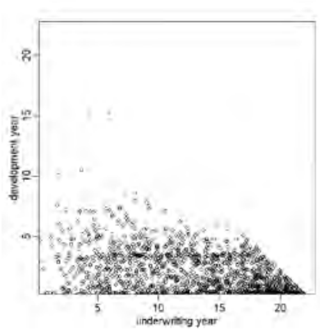
- We start with the problem of predicting number of claims.
- The aim is to **estimate 2-dimensional density, which is only observed in a triangle**.
- Thus, we have a density estimation/forecasting problem.
- Outstanding liabilities consist of **integrals of such a density.**

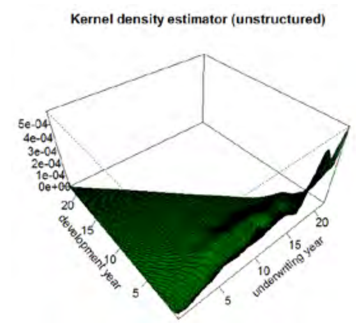
Specifications:

1. The data are arranged in a two dimensional space: **still a triangle.**
2. But the **time is continuous.**
3. The dependencies in the data are modelled as **time effects**: underwriting, development,…

Solving the problem in two steps

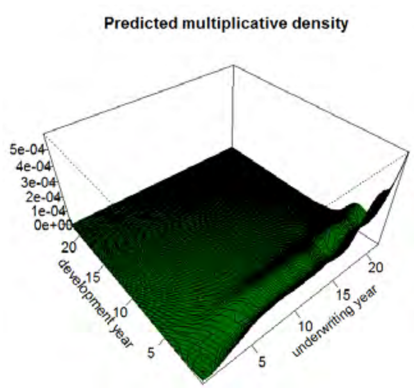
1. Density estimation with a triangular support



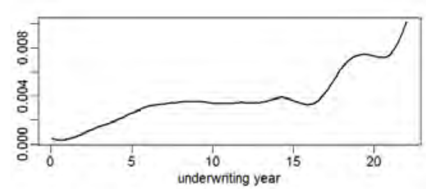


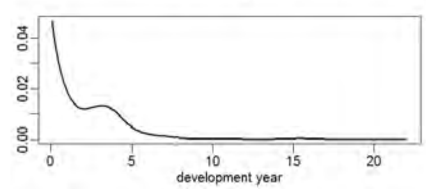
Look for **the best density estimator available**

2. Forecasting problem: the density in the whole square



$$f(x, y) = f_1(x) f_2(y)$$



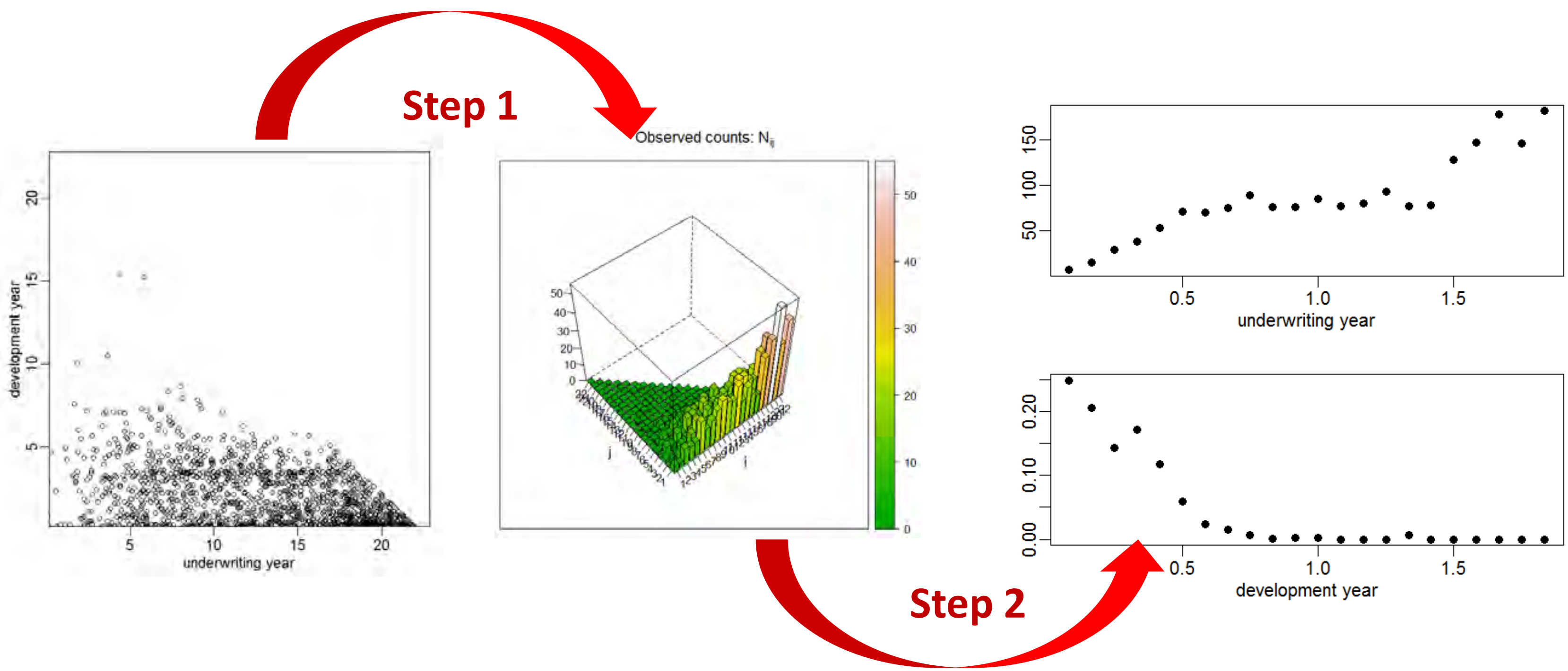


Start with a simpler model: **multiplicative structure**

Cass Business School
CITY UNIVERSITY LONDON

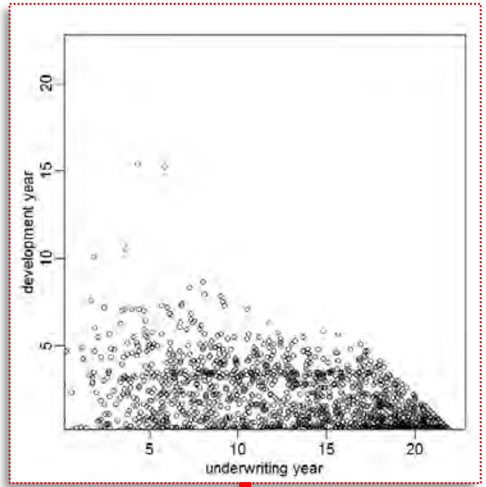
Reformulating classical chain ladder in this framework

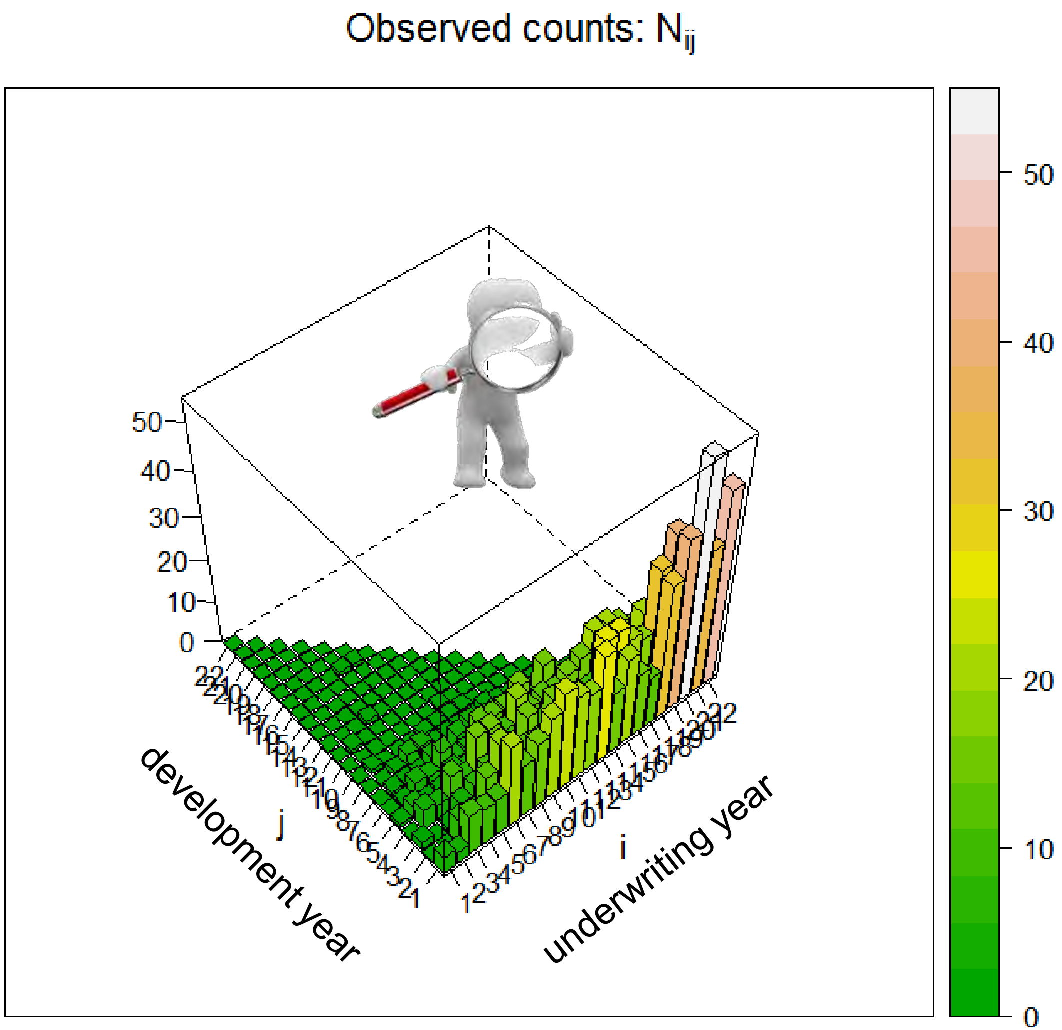
Chain ladder starts from a **histogram** of the granular data. Then this histogram is projected on a **multiplicative structure** for forecasting the future



Cass Business School
CITY UNIVERSITY LONDON

What we can learn… 1. Chain ladder is indeed granular!

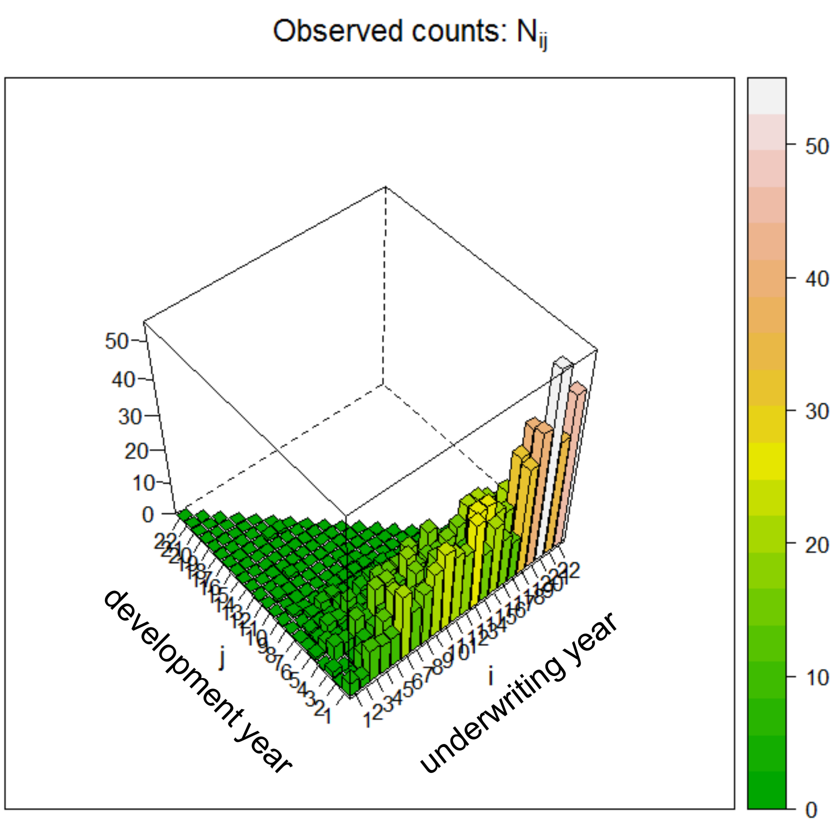


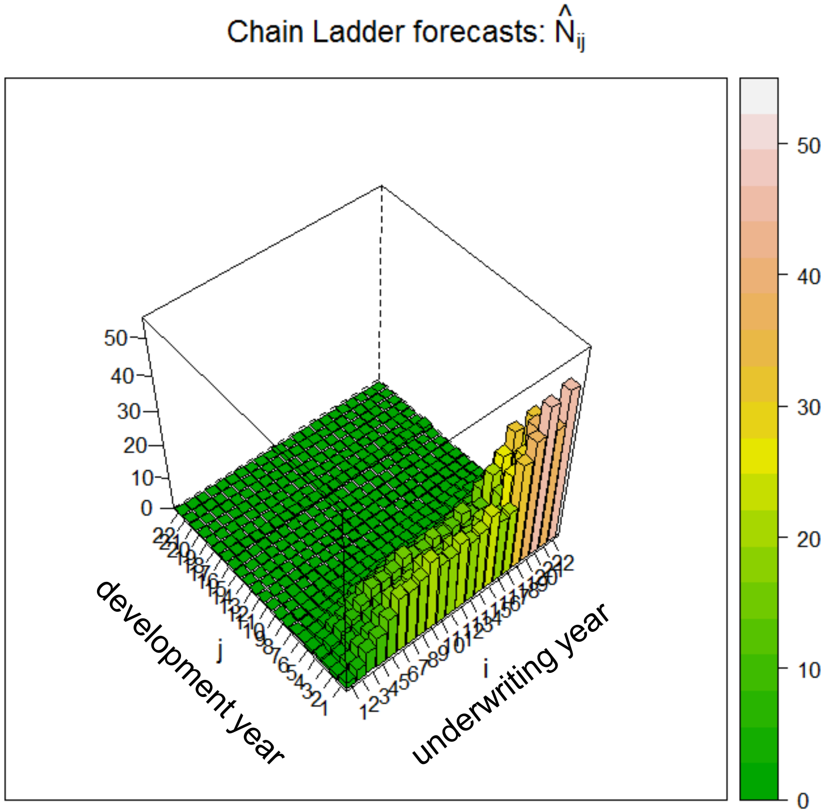


A **histogram** is a common graphical tool to explore, and show, the underlying distribution of a set of **continuous data.**

It is, maybe, the **simplest density estimator.**

What we can learn… 2. The multiplicative structure





Future counts are predicted assuming: $\widehat{N}_{ij} = \widehat{\alpha}_i \widehat{\beta}_j$

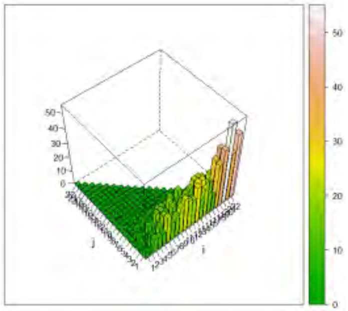
Let's see it in more detail…

Cass Business School
CITY UNIVERSITY LONDON

The classical Poisson model for chain ladder

We consider an incremental triangle

$$\aleph_k = \{N_{ij},\ (i,j) \in \mathcal{I}\}$$

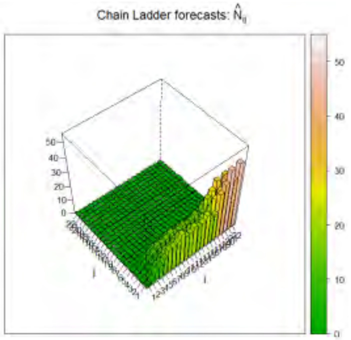


Assume that N_{ij} ‘s are independently Poisson distributed with means having a **multiplicative parameterization**:

$$\mathrm{E}[\mathrm{N_{ij}}] = \alpha_\mathrm{i}\beta_\mathrm{j}.$$

Then calculate the maximum likelihood estimates of the parameters: $\{\widehat{\alpha}_i, \widehat{\beta}_j : i, j+1 = 1, \ldots, k\}$

Predict the future entries as: $\widehat{N}_{ij} = \widehat{\alpha}_i \widehat{\beta}_j$



Cass Business School
CITY UNIVERSITY LONDON

The classical Poisson model for chain ladder

The log-likelihood:

$$l\left(\alpha_i, \beta_j; \aleph_m\right) = \sum_{i=1}^{m} \sum_{j=1}^{m-i+1} \left\{-\alpha_i\beta_j + N_{ij}\log(\alpha_i\beta_j) - \log(N_{ij}!)\right\}$$

The score:

$$\frac{\partial l\left(\alpha_i, \beta_j; \aleph_m\right)}{\partial \alpha_i} = \sum_{j=1}^{m-i+1} \left\{-\beta_j + \frac{N_{ij}}{\alpha_i}\right\} = 0 \quad \Rightarrow \quad \alpha_i = \frac{\sum_{j=1}^{m-i+1} N_{ij}}{\sum_{j=1}^{m-i+1} \beta_j}$$

$$\frac{\partial l\left(\alpha_i, \beta_j; \aleph_m\right)}{\partial \beta_j} = \sum_{i=1}^{m-j+1} \left\{-\alpha_i + \frac{N_{ij}}{\beta_j}\right\} = 0 \quad \Rightarrow \quad \beta_j = \frac{\sum_{i=1}^{m-j+1} N_{ij}}{\sum_{i=1}^{m-j+1} \alpha_i}$$

A constraint is necessary to identify the **parameters**

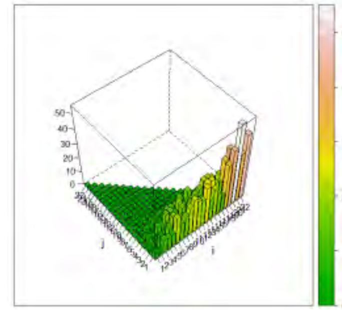
$$\sum_{j=1}^{m} \beta_j = 1$$

Explicit expressions for the estimates (e.g. Verrall 1991)

Cass Business School
CITY UNIVERSITY LONDON

Reformulating the classical approach

Consider a **histogram estimator** of the density in the triangle:

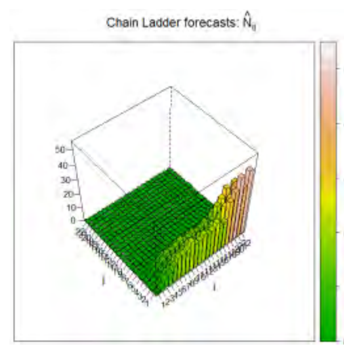


$$\widehat{f}_{hist}(x,y) = \frac{N_{ij}}{n\Lambda_m^2}, \qquad (x,y) \in B_{ij} = (i, i+1] \times (j, j+1]$$
$$n = \sum N_{ij}, \;\; \Lambda_m \equiv \text{bin length}$$

It is known that:

$$N_{ij} \hookrightarrow B(n, p_{ij}) \approx P(np_{ij}), \;\; p_{ij} = \int_{B_{ij}} f(x,y)dxdy$$

which justifies a GLM model with Poisson error distribution for the future counts.



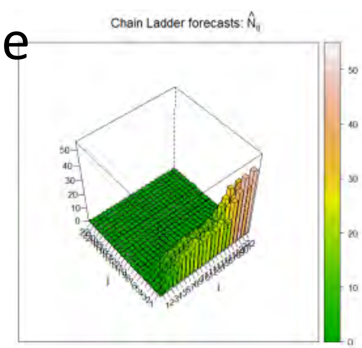
Thus, **the density problem is treated as a regression problem on an aggregated base** (Fan and Gijbels 1995}:

$$N_{ij} = r(i,j) + \varepsilon_{ij} \quad \Longrightarrow \quad f(x,y) = \frac{r(x,y)}{n\Lambda_m^2}$$

Reformulating the classical approach

Assume a **multiplicative** and **parametric** structure for the regression function:

$$r(x,y) = \alpha_i \beta_j \quad \text{for } (x,y) \in B_{ij}$$



This implies that the target density

$$f(x,y) = f_1(x) f_2(x) \quad \text{for } (x,y) \in B_{ij}$$

$$\text{with } \begin{cases} f_1(x) = \kappa_1 \alpha_i \\ f_2(y) = \kappa_2 \beta_j \end{cases} \quad \text{(piece-wise constant densities)}$$

From these assumptions, the equations to derive the estimates can be rewritten as:

$$\alpha_i = \frac{\sum_{j=1}^{m-i+1} N_{ij}}{\sum_{j=1}^{m-i+1} \beta_j} \quad \Rightarrow \quad f_1(x) = \frac{\int_{\mathcal{I}_x} \widehat{f}_{hist}(x,y)dx}{\int_{\mathcal{I}_x} f_2(y)dy}$$

$$\beta_j = \frac{\sum_{i=1}^{m-j+1} N_{ij}}{\sum_{i=1}^{m-j+1} \alpha_i} \quad \Rightarrow \quad f_2(x) = \frac{\int_{\mathcal{I}_y} \widehat{f}_{hist}(x,y)dx}{\int_{\mathcal{I}_y} f_1(x)dx}$$

$$\begin{cases} \mathcal{I}_x = \{y | (x,y) \in \text{triangle}\} \\ \mathcal{I}_y = \{x | (x,y) \in \text{triangle}\} \end{cases}$$

Cass Business School
CITY UNIVERSITY LONDON

Summary:

- Classical chain ladder is indeed a granular method, because the histogram works on continuous data.
- Assumptions for forecasting the target density in the future:
  1. A **multiplicative structure** for the 2-dimensional density.
  $$f(x,y) = f_1(x)f_2(x)$$
  2. The densities in the underwriting and development directions are **piece-wise constant.**

Advantages of this approach: simplicity, the problem can be treated as a parametric problem with maximum likelihood solutions.

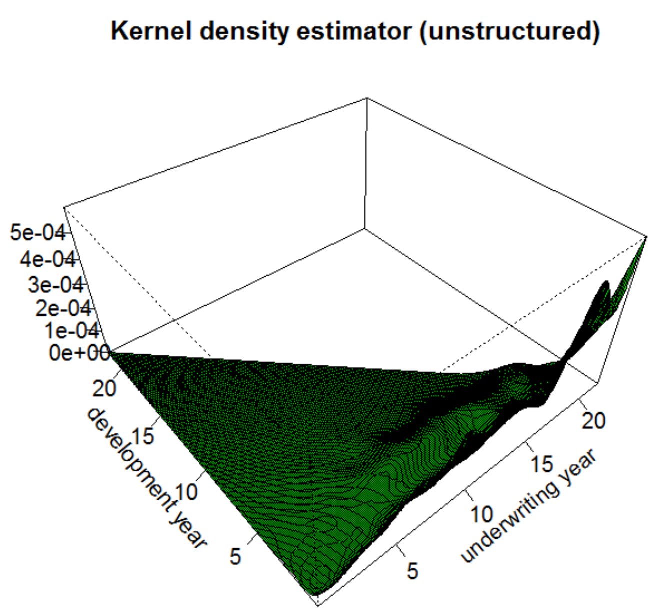
Drawbacks:

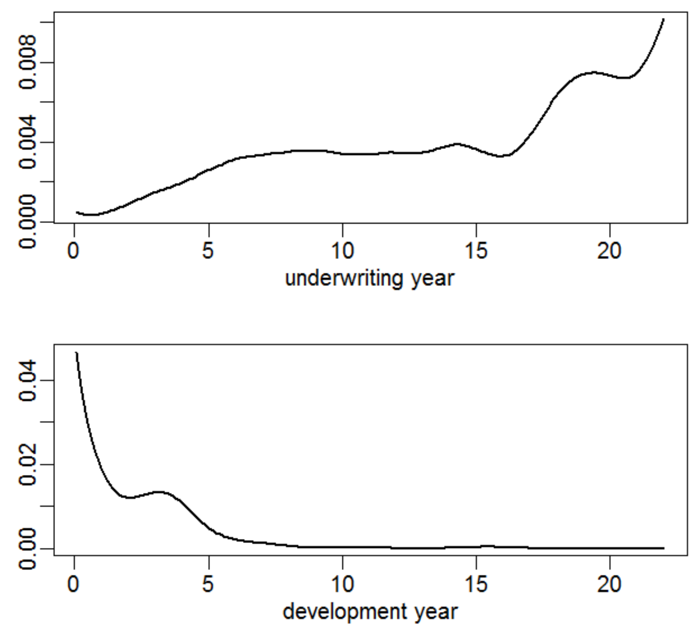
- The **histogram is an inefficient estimator of the density.**
- It leads to **discrete time effects.**

Continuous Chain Ladder: the natural improvement

1. **Replace the histogram by a kernel estimator** of the density: the natural way to improve on histograms

2. Assume a **multiplicative structure** but with **non-parametric time effects** (continuous densities)



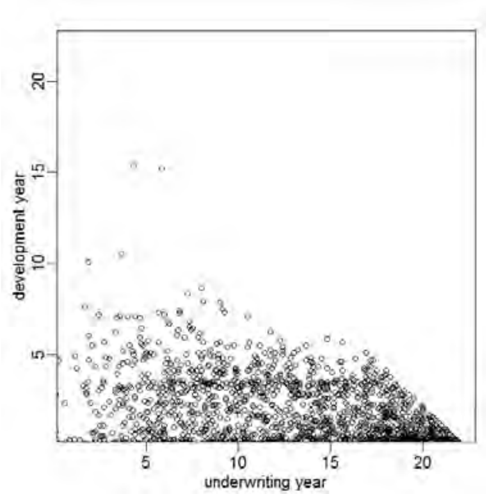


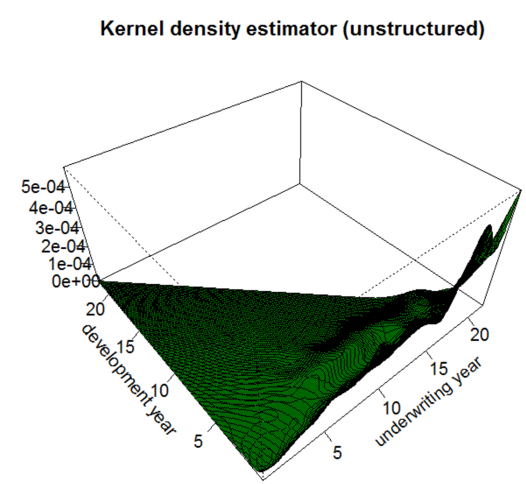
Cass Business School
CITY UNIVERSITY LONDON

Continuous Chain Ladder: mathematical formulation

1. Density estimation with a triangular support:

Let $\{Z_1 = (X_1, Y_1)^t, \ldots, Z_n = (X_n, Y_n)^t\}$ be an i.i.d. random sample from a population $Z = (X, Y)^t$, having two-dimensional continuous density f 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.





Continuous Chain Ladder: mathematical formulation

Unstructured kernel estimator of the density:

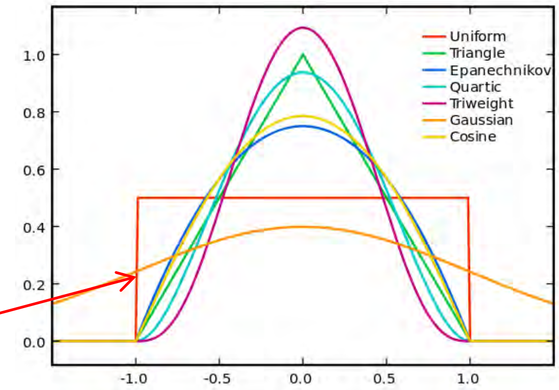
For any $z_0 = (x_0, y_0)^t \in \mathcal{I}$, the simpler kernel density estimator is:

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

- $\mathcal{K}_h(x, y) = K_{h_1}(x) K_{h_2}(y)$ is a multiplicative **two-dimensional kernel**, with $K_{h_1}(x) = h_1^{-1} K(x/h_1)$, $K_{h_2}(y) = h_2^{-1} K(y/h_2)$ and K being a unidimensional density function,
- $h = (h_1, h_2)^t \in \mathbb{R}_+^2$ is the **bandwidth**, with $|h| = h_1 h_2$.

Relation with the histogram density estimator:

- Histograms consider the bandwidth as the bin length divided by 2.
- The kernel defines the weight-assigning function. Histograms use a uniform or rectangular kernel.



Cass Business School
CITY UNIVERSITY LONDON

Continuous Chain Ladder: mathematical formulation

Improved kernel density estimators:

Boundary correction

- The **local linear** (LL) density estimator (Nielsen 1999), which is defined at each $z_0 = (x_0, y_0)^t \in \mathcal{I}$ as the solution $\widehat{f}^{\mathcal{I}}_{LL,h}(z_0) = \widehat{\Theta}_0$ of the following minimization problem:

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

where $\widetilde{f}_b(z) = n^{-1}(b_1 b_2)^{-1} \sum_{i=1}^n \mathcal{K}_b(z - z_0)$.

Bias correction

- The **multiplicative bias corrected** (MBC) density estimator, which is defined:

$$\widehat{f}^{\mathcal{I}}_{MBC,h}(z_0) = \widehat{f}^{\mathcal{I}}_{LL,h}(z_0) \widehat{g}^{\mathcal{I}}_{LL,h}(z_0)$$

with $\widehat{g}^{\mathcal{I}}_{LL,h}(\cdot)$ being the local linear estimator of the ratio $f(\cdot)/\widehat{f}^{\mathcal{I}}_{LL,h}(\cdot)$.

Cass Business School
CITY UNIVERSITY LONDON

Continuous Chain Ladder: mathematical formulation

2. Forecasting problem: the density on the whole square

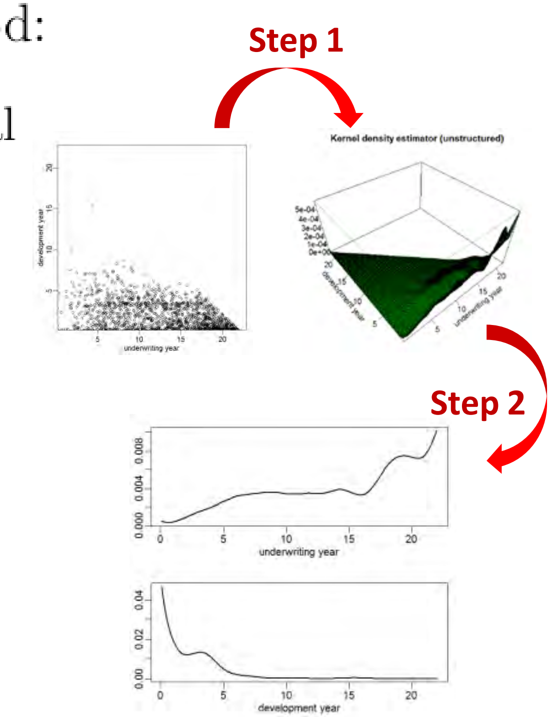
Assume that the density in $\mathcal{S} = \{z = (x, y)^t | 0 \le x, y \le T\}$ is multiplicative i.e. $f(x, y) = f_1(x) f_2(y)$. An extension of the marginal integration method (Linton and Nielsen 1995) provide estimators of f_1 and f_2 through the following two-step method:

Step 1. From the available data estimate the two-dimensional density in the observation set $\mathcal{I}$ by an estimator $\widehat{f_h^{\mathcal{I}}}(x, y)$ (LL or MBC).

Step 2. Assume $f(x, y) = f_1(x) f_2(y)$ and estimate f_1 and f_2 through the following minimization:

$$\min_{f_1, f_2} \int_{\mathcal{I}} \left(\widehat{f_h^{\mathcal{I}}}(x, y) - f_1(x) f_2(y) \right)^2 w(x, y) dx dy,$$

with $w(x, y)$ being a weighting function.



Continuous Chain Ladder: mathematical formulation

An algorithm to perform Step 2:

(i) Consider an initial estimator of f_1, $\widehat{f}_1^{(0)}$.

(ii) Estimate the density f_2 by

$$\widehat{f}_2^{(1)}(y) = \frac{\int_{\mathcal{I}_y} \widehat{f}_h^{\mathcal{I}}(x,y)dx}{\int_{\mathcal{I}_y} \widehat{f}_1^{(0)}(x)dx},$$

with $\mathcal{I}_y = \{x|(x,y) \in \mathcal{I}\}$.

(iii) Using $\widehat{f}_2^{(1)}$, calculate the updated estimator for f_1 by

$$\widehat{f}_1^{(1)}(x) = \frac{\int_{\mathcal{I}_x} \widehat{f}_h^{\mathcal{I}}(x,y)dy}{\int_{\mathcal{I}_x} \widehat{f}_2^{(1)}(y)dy}$$

with $\mathcal{I}_x = \{y|(x,y) \in \mathcal{I}\}$.

(iv) Repeat steps (ii)-(iii) until convergence.

Reminder of the chain ladder equations:

$$f_2(x) = \frac{\int_{\mathcal{I}_y} \widehat{f}_{hist}(x,y)dx}{\int_{\mathcal{I}_y} f_1(x)dx}$$

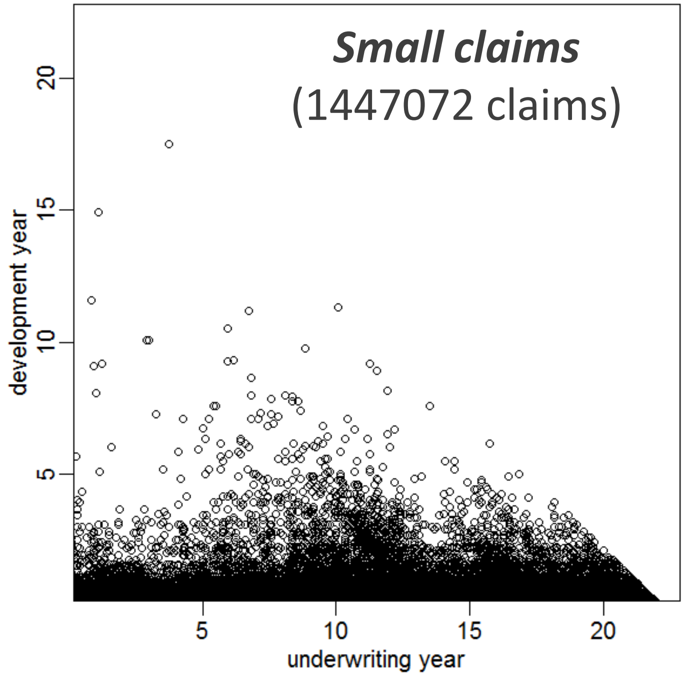
$$\beta_j = \frac{\sum_{i=1}^{m-j+1} N_{ij}}{\sum_{i=1}^{m-j+1} \alpha_i}$$

$$f_1(x) = \frac{\int_{\mathcal{I}_x} \widehat{f}_{hist}(x,y)dx}{\int_{\mathcal{I}_x} f_2(y)dy}$$

$$\alpha_i = \frac{\sum_{j=1}^{m-i+1} N_{ij}}{\sum_{j=1}^{m-i+1} \beta_j}$$

Illustration: Prediction of the outstanding number of claims

We consider two data sets provided by a major insurer on a monthly base. The data are the number of reported claims, and it has been arranged in a triangle where the **development period** corresponds with the **reporting period**.



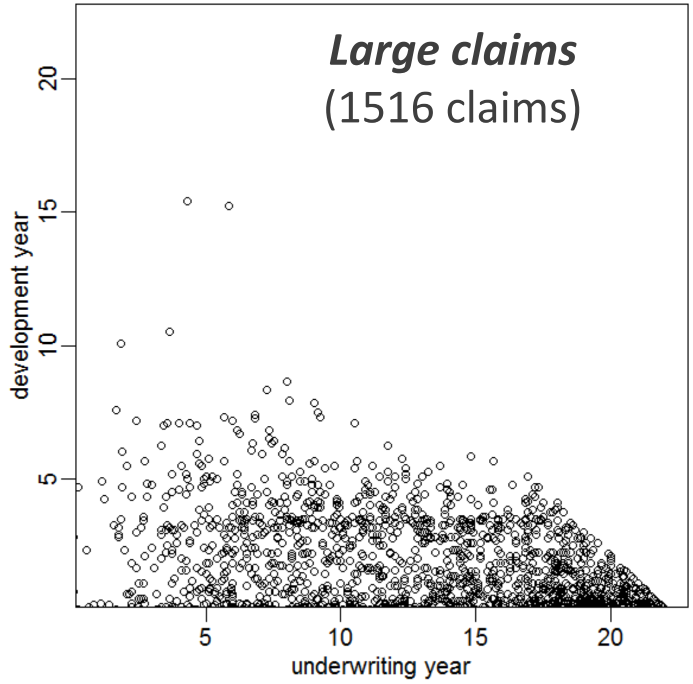


Illustration: Comparing four methods to solve the problem

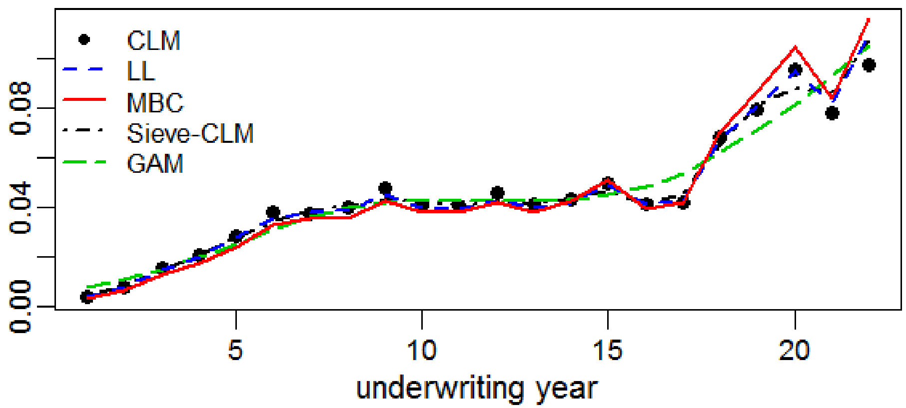
- ✓ Classical **Chain Ladder** from a yearly run-off triangle.
- ✓ Two versions of **Continuous Chain Ladder** with two kernel unstructured density estimators (LL and MBC).
- ✓ **GAM method of England and Verrall (2001)**: starting from the histogram the time effects are estimated using smoothing splines

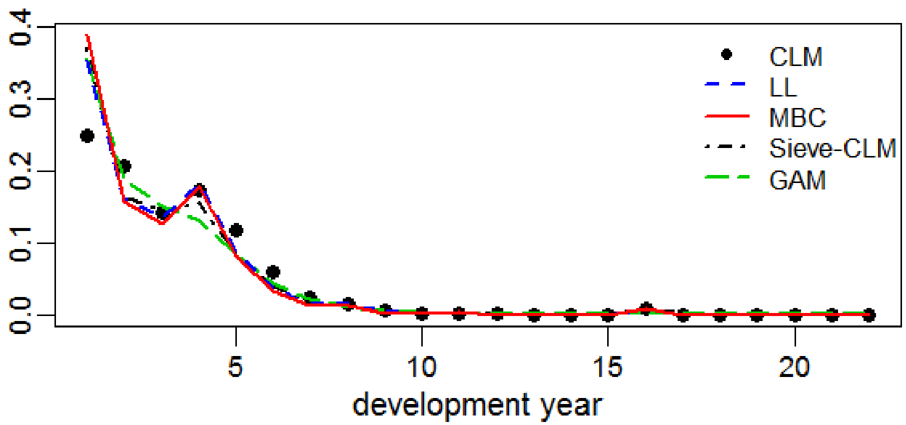
$$\log\left(N_{ij}\right) = s_{\theta_i}(i) + s_{\theta_j}(j) + \varepsilon_{ij}$$

- ✓ A **sieve method on monthly chain ladder parameters**: providing smoothed chain ladder time effects using local regression.

Illustration: results for large claims

Estimated time effects





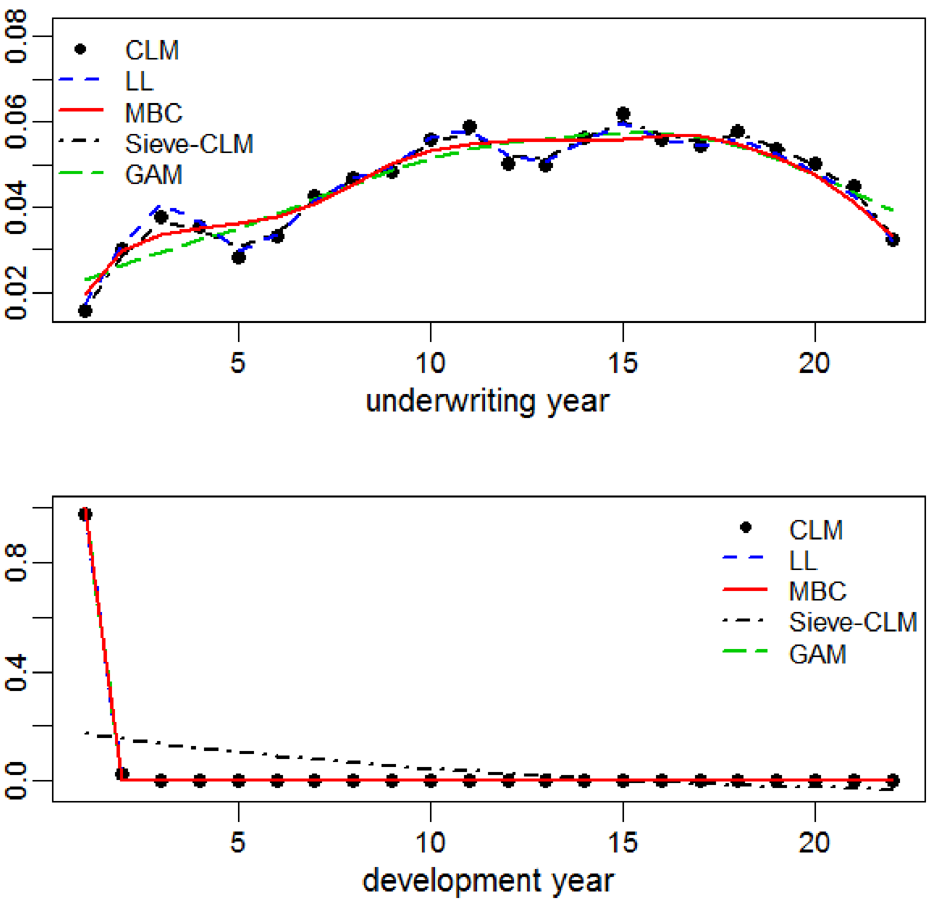
Predictions for future calendar years

Future	CLM	LL	MBC	Sieve-CLM	GAM
1	118	132	104	133	127
2	86	93	71	93	88
3	65	72	54	70	64
4	38	43	32	43	40
5	19	21	15	21	22
6	9	10	7	10	11
7	5	6	4	6	6
8	2	3	2	3	3
9	1	1	1	1	2
10	1	1	1	1	1
11	1	1	1	1	1
12	1	1	1	1	1
13	1	1	1	1	0
14	1	1	1	1	0
15	1	1	1	1	0
16	0	1	0	1	0
17	0	0	0	0	0
Total	350	390	295	388	367

Cass Business School
CITY UNIVERSITY LONDON

Illustration: results for small claims

Estimated time effects



Predictions for future calendar years

Future	CLM	LL	MBC	Sieve-CLM	GAM
1	1,312	1,105	1,065	80,418	1,303
2	98	95	92	65,584	46
3	42	40	37	52,615	33
4	19	18	16	40,921	22
5	9	9	8	30,262	13
6	6	6	5	20,956	7
7	4	4	3	13,167	4
8	2	2	2	6,717	3
9	2	2	1	1,433	2
10	1	1	1	-2,583	1
11	1	1	1	-5,575	1
12	1	1	1	-8,009	0
13	1	0	1	-9,745	0
14	1	0	1	-10,524	0
15	0	0	0	-10,443	0
16	0	0	0	-9,920	0
17	0	0	0	-8,933	0
Total	1,501	1,284	1,235	228,145	1,435

Cass Business School
CITY UNIVERSITY LONDON

Illustration: testing results against experience

The validation strategy:

1. Cut c=1,2,…,5 diagonals (years) from the observed triangle.
2. Apply the four estimation methods.
3. Compare forecasts and actual values.

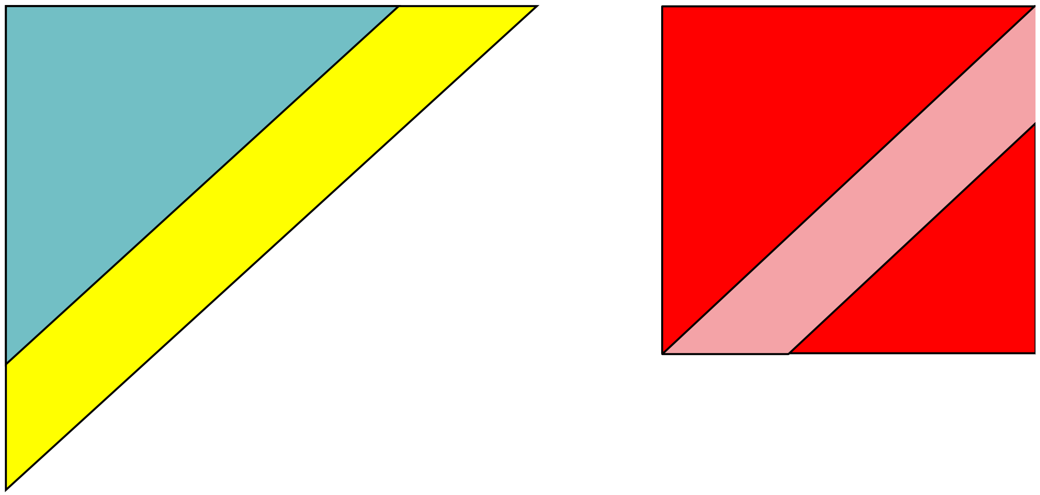
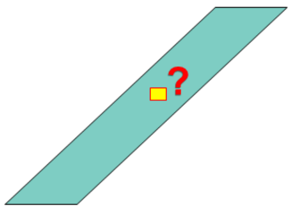


Illustration: testing results against experience

Three possible objectives:

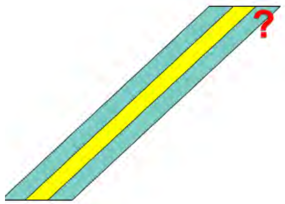
1. Predictions of the individual cells

$$Rerr_1^c = \frac{\sum_{(i,j)\in\tilde{\mathcal{J}}_c}(\widehat{N}_{ij} - N_{ij})^2}{\sum_{(i,j)\in\tilde{\mathcal{J}}_c} N_{ij}^2}$$



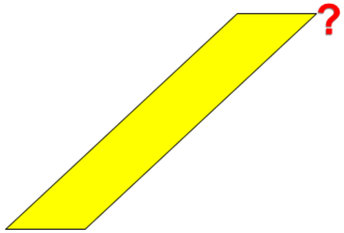
2. Predictions by calendar years

$$Rerr_2^c = \frac{\sum_{k=1}^{c}(\widehat{D}_{k;c} - D_{k;c})^2}{\sum_{k=1}^{c}(D_{k;c})^2}$$



3. The prediction of the overall total

$$Rerr_3^c = \frac{|\widehat{R}_c - R_c|}{R_c}$$



Cass Business School
CITY UNIVERSITY LONDON

Illustration: testing results against experience

Large claims

Objective	c	LL	MBC	Sieve-CLM	GAM
Cells	1	1.11	0.82	1.23	1.06
	2	1.02	0.59	1.22	1.03
	3	1.11	0.84	1.15	0.76
	4	1.13	0.94	1.33	0.86
	5	1.03	0.99	0.88	0.79
Calendar	1	1.09	0.34	1.73	1.72
	2	1.05	0.56	1.45	1.02
	3	1.12	0.66	1.76	0.93
	4	1.42	0.82	2.57	0.96
	5	1.32	0.85	0.89	1.47
Total	1	1.09	0.34	1.73	1.72
	2	1.05	0.54	1.46	1.03
	3	1.13	0.30	1.95	1.00
	4	1.60	0.26	3.07	1.04
	5	1.32	0.21	0.89	1.46

Small claims

Objective	c	LL	MBC	GAM
Cells	1	0.77	0.43	0.81
	2	0.83	0.54	0.77
	3	0.67	0.50	0.71
	4	0.75	0.49	0.67
	5	0.70	0.96	0.53
Calendar	1	0.78	0.46	0.69
	2	0.85	0.58	0.69
	3	0.65	0.46	0.53
	4	0.76	0.52	0.61
	5	0.70	0.91	0.40
Total	1	0.78	0.46	0.69
	2	0.86	0.60	0.62
	3	0.69	0.52	0.53
	4	0.77	0.52	0.52
	5	0.76	0.27	0.15

Relative errors with respect to the classical chain ladder method
(values lower than 1 indicate an improvement on chain ladder)

Cass Business School
CITY UNIVERSITY LONDON

Conclusions

- This work establishes a link between classical chain ladder and modern mathematical statistics.
- The interpretation of classical chain ladder as a structured histogram estimator has a number of immediate implications for further developments.
- “Continuous Chain Ladder” is the natural kernel smoother improving the histogram of classical chain ladder.
- Further work is required to refine and extend the methods.
  - Bandwidth selection is not immediate in this framework.
  - What happens with paid data??...

Cass Business School
CITY UNIVERSITY LONDON

When the aim is forecasting payments

1. Continuous chain ladder as well as classical chain ladder is a model for counts data: the goal is just a density.
2. With paid data, the outstanding liabilities are weighted integrals of a density. Now, a severity function plays a role in the model.
3. But also the dependencies in the data should be considered through a more complicated model.
4. But the way is almost prepared for the extension: **Continuous Double Chain Ladder** (**CCL+DCL = CDCL**)

Double Chain Ladder model (Martínez-Miranda, Nielsen and Verrall 2012)