A hierarchical model for micro-level stochastic loss reserving

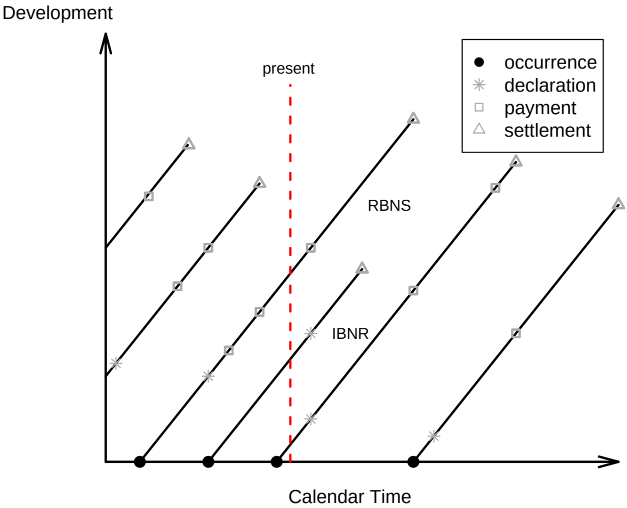
joint work with K. Antonio[1] and E.W. Frees[2]

44th Actuarial Research Conference
Madison, Wisconsin 30 Jul - 1 Aug 2009

E.A. Valdez
University of Connecticut
Storrs, Connecticut

[1]U. of Amsterdam
[2]U. of Wisconsin – Madison

Dynamics of claims reserving



Synopsis

- Put focus on RBNS claims: *Reported But Not Settled*.
- Use **micro-level** data to predict future development of open claims.
- Develop a **hierarchical** model.
- *"A hierarchical model for micro–level stochastic loss reserving."*

The data

- Data are from the *General Insurance Association of Singapore*.
- Observations are from one company over 10-year period: Jan 1993 – Jul 2002.

  ⇒ "present moment" in this case–study is 25 Jul 2002.
- **Policy file**: characteristics of policyholder and vehicle insured

  ⇒ age, gender, vehicle type, vehicle age, . . .
- **Claims file**: keeps track of each accident claim filed with the insurer

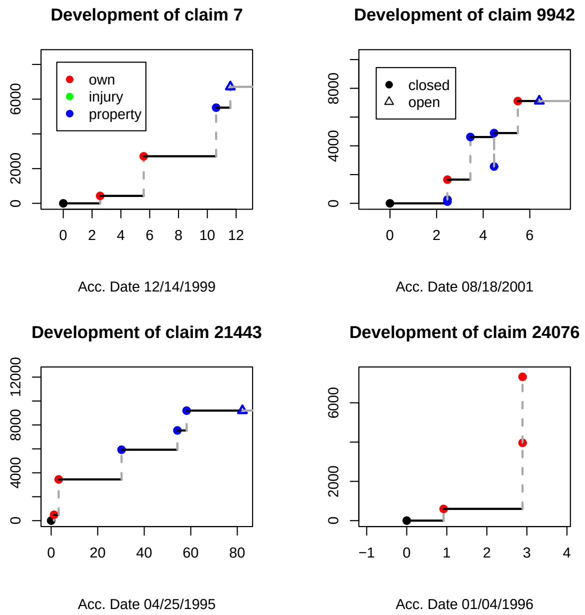
  ⇒ linked to policy file, contains accident date.
- **Payments file**: reports each payment made during observation period.

  ⇒ linked to claims file, with payment date, size and type.

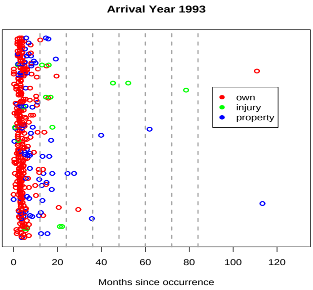
The data

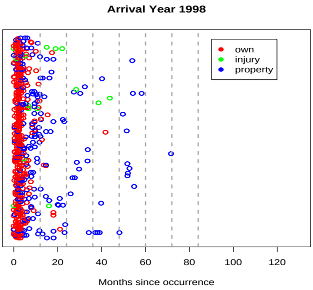
- A claim will have **multiple payments** during its run–off.
- **Payment types** may be:
  - own damage (*O*) (including injury, property, fire, theft);
  - injury (*I*) to a party other than the insured;
  - property damage (*P*).
- Combinations of these types may also occur.
- Frees and Valdez (2008, JASA) summarized the many payments per claim into one single claim amount.

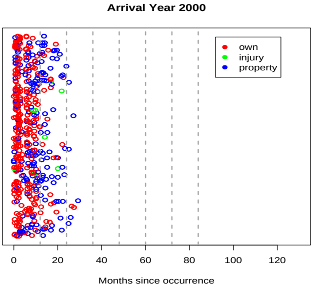
The data


Development of claim 7
own
injury
property
0
2000
6000
0
2
4
6
8
10
12
Acc. Date 12/14/1999
Development of claim 9942
closed
open
0
4000
8000
0
2
4
6
Acc. Date 08/18/2001
Development of claim 21443
0
4000
8000
12000
0
20
40
60
80
Acc. Date 04/25/1995
Development of claim 24076
0
2000
6000
−1
0
1
2
3
4
Acc. Date 01/04/1996

The data


Arrival Year 1993
own
injury
property
0
20
40
60
80
100
120
Months since occurrence


Arrival Year 1998
own
injury
property
0
20
40
60
80
100
120
Months since occurrence


Arrival Year 2000
own
injury
property
0
20
40
60
80
100
120
Months since occurrence

A traditional actuarial display

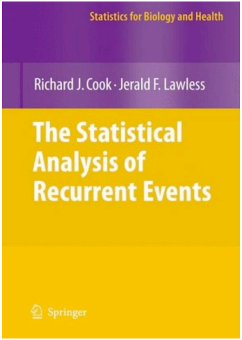
- **Run–off triangle**: aggregate claims per arrival year (AY) and development year (DY) combination.
- Run–off triangle for property (P) payments: (in ’000s, non–cumulative)

Arrival Year	Development Year 1	2	3	4	5	6	7	8	9	10
1993	205.3	847.6	226.3	77.9	47.9	40.6	10.2	1.8	0.0	0.6
1994	1,081.3	1,750.4	534.7	153.8	73.0	51.1	16.2	37.3	5.8	
1995	900.9	1,822.7	578.5	202.0	54.1	48.2	9.5	1.3		
1996	1,272.8	1,816.9	583.7	255.2	44.2	24.1	11.4			
1997	1,188.7	2,257.9	695.2	166.8	92.1	12.9				
1998	1,235.4	3,250.0	649.9	211.2	74.1					
1999	2,209.8	3,718.7	818.4	266.3						
2000	2,662.5	3,487.0	762.7							
2001	2,457.3	3,650.3								
2002	673.7									

- Common statistical techniques: chain–ladder, distributional, Bayesian, GLMs, . . .
- Modeling individual claims run-off is less developed in the literature.

Micro–level data: literature

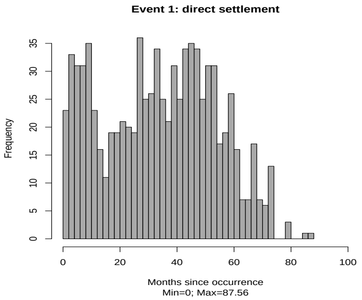
- **Suggestions** from actuarial literature: England and Verrall (2002), Taylor and Campbell (2002), Taylor, McGuire, and Sullivan (2006).
- Some **actuarial** papers:
  - Arjas (1989, ASTIN), Norberg (1993, ASTIN), Norberg (1999, ASTIN);
  - Haastrup and Arjas (1996, ASTIN);
  - Larsen (2007, ASTIN);
  - Zhao, Zhou, and Wang (2009, IME).
- **Statistical** resource: Cook and Lawless (2007), *Statistical analysis of recurrent events*.

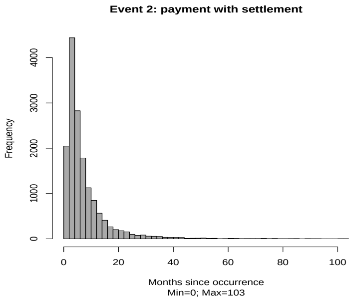


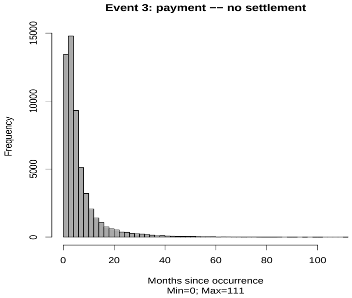
Observable data structure

- total number of claims in the data set is $n = 43,729$;
- N_i, number of “events” in development period of claim i;
- T_{ij}, **time of event** j, in months since the accident date ($T_{i0} = 0$ is accident date and T_{iN_i} is settlement date);
- C_i time of **censoring**;
- E_{ij} **type of event** j. We distinguish:
  - event type 1: direct settlement without any payments;
  - event type 2: payment with settlement;
  - event type 3: payment without settlement.
- M_{ij} **type of payment** for event j of claim i.
- P_{ijk} **size of payment** of type k (k being ‘own damage’ (O), ‘injury’ (I) or ‘property’ (P)) for event j of claim i.

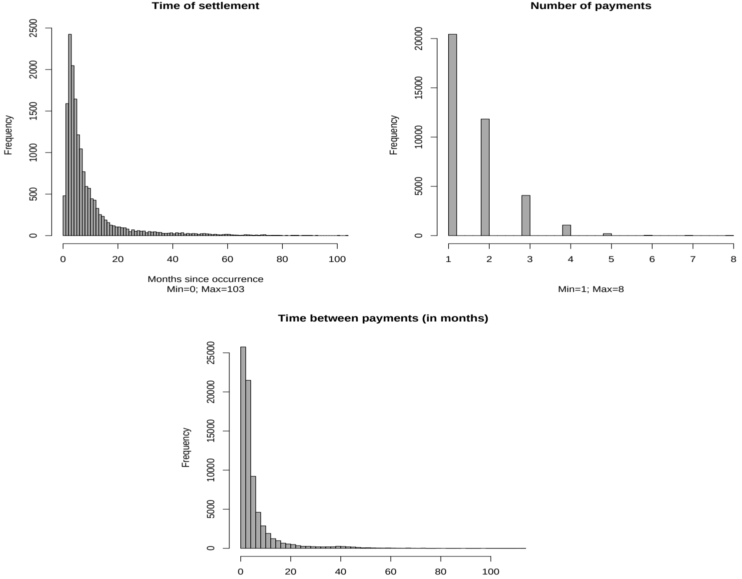
Timing of events, per event type







Time of settlement, number of payments, times between payments

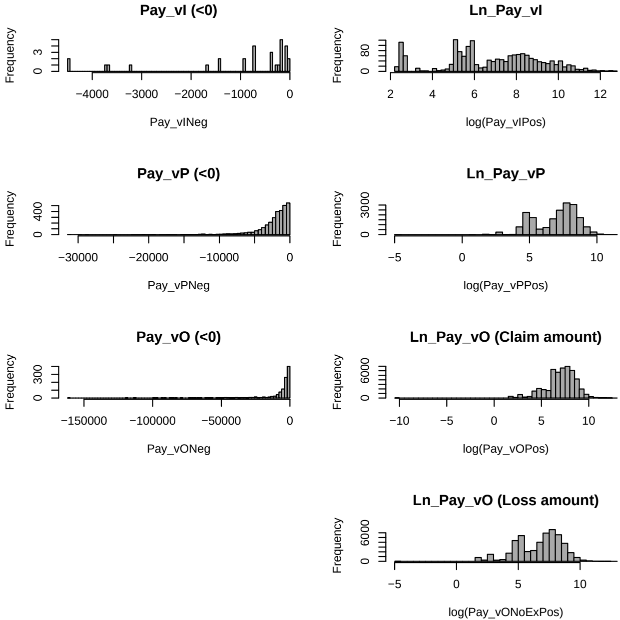


Payment types

- Number of payments per type:

	Claim Type			
	(I)	(O)	(P)	
Number	1,417 (1.95%)	45,950 (63.3%)	21,775 (30%)	
	(I,O)	(I,P)	(O,P)	(O,I,P)
Number	107 (0.147%)	319 (0.439%)	3017 (4.16%)	9 (0.012%)

Distribution of payments



Model formulation

- A **claim** i ($i = 1, \ldots, n_c$) is a combination of
  - accident date ('$\mathbf{AD}_i$');
  - set of covariates $\mathbf{C}_i$;
  - development process $\mathbf{X}_i$:
    $\mathbf{X}_i = (\{E_i(v), M_i(v), \mathbf{P}_i(v)\})_{v \in [0, T_{iN_i}]}$;
- **Development process** $\mathbf{X}_i$ is a jump process. 3 building blocks are used:
  - $E_i(t_{ij}) := E_{ij}$ is the type of the jth event in the development of claim i, occurring at time t_{ij};
  - If this event includes a payment, its payment is given by $M_i(t_{ij}) := M_{ij}$;
  - Corresponding payment vector is $\mathbf{P}_{ij}$.

Time to events

- **Intensity modeling** with single type of events at times t_{ij}:

$$L_i \quad = \quad \left(\prod_{j=1}^{N_i} \lambda_i(t_{ij})\right) \exp\left(-\int_0^{\tau_i} \lambda_i(u)du\right).$$

  - $[0, \tau_i]$ is the period of observation of subject i with $\tau_i = \min\left(T_{iN_i}, C_i\right)$.
  - $\lambda_i(t)$ is the **event intensity** (or hazard rate) at time t for subject i.

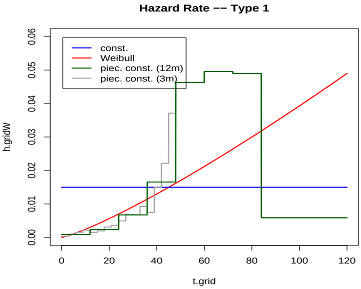
- For **multitype events**: each "subject" is at risk of m different types of recurrent events.
  - Specify intensity function for each type of event ($k = 1, \ldots, m$) with $\lambda_{ik}(t)$.

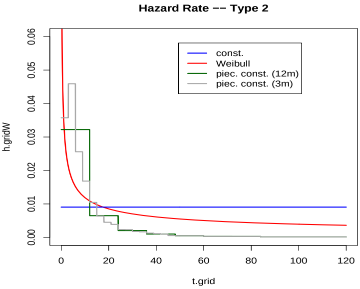
Time to events

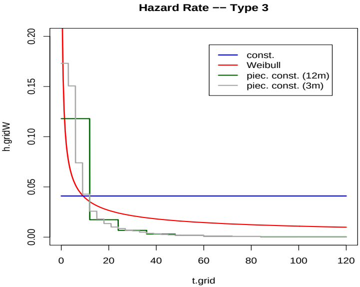
- **How to specify the intensity functions** $\lambda_1(t)$ (for event 1), $\lambda_2(t)$ (for event 2) and $\lambda_3(t)$ (for event 3)?
- Techniques from survival analysis: $(k = 1, 2, 3)$
  - exponential: $\lambda_k(t) := \lambda_k$;
  - Weibull: $\lambda_k(t) := \alpha_k \gamma_k t^{\alpha_k - 1} e^{-\gamma_k t^{\alpha_k}}$;
  - Cox model: $\lambda_k(t) := \lambda_{0k}(t) \exp(\boldsymbol{z}_k' \boldsymbol{\beta}_k)$;
  - piecewise constant:

$$\lambda_k(t) = \begin{cases} \lambda_{k1} \text{ for } 0 \leq t < t_{k1} \\ \lambda_{k2} \text{ for } t_{k1} \leq t < t_{k2} \\ \vdots \\ \lambda_{kd} \text{ for } t_{kd-1} \leq t < t_{kd}. \end{cases}$$

Hazard rates per event type







Payment type

- M_{ij} represents the combination of payments observed at t_{ij}.
- 7 combinations are possible: I, O, P, (I, O), (I, P), (O, P) and (O, I, P).
- Claim type is modeled with **multinomial logit** model:

$$Pr(M_{ij} = m_{ij}) \quad = \quad \frac{\exp V_{ij,m}}{\sum_{s=1}^{7} \exp\left(V_{ij,s}\right)},$$

with $V_{ij,m} = \boldsymbol{x}_{ij}' \beta_{M,m}$.

- Covariate information used in multinomial model:
  - **Type of vehicle**, vehicle age, age of driver;
  - Arrival Year, **Development Year**.

Payments

- Given M_{ij} for the event at time t_{ij}, $\boldsymbol{P}_{ij}$ gives corresponding severities.
- For the **sign of a payment**, use:

$$I_{ijk} = \begin{cases} 1 \text{ if } P_{ijk} > 0 \\ 0 \text{ if } P_{ijk} < 0, \end{cases}$$

  and $s_{ijk} = Pr(I_{ijk} = 1)$.
- Use **logistic regression** to model the sign of P_{ijk}:

$$\text{logit}(s_{ijk}) = \boldsymbol{x}'_{ij}\beta_{S,k}.$$

- Covariate information used in logistic models:
  - Development year;
  - Number of previous injury/own damage/property payments.

Negative part of payments

- **Burr regression**:

$$f_P(p) = \frac{\lambda\beta^{\lambda}\tau p^{\tau-1}}{(\beta+p^{\tau})^{\lambda+1}},$$

with $\tau_{ijk} = \exp\left(\boldsymbol{x}_{ijk}^{'}\boldsymbol{\beta}_{P,k}\right)$ with k for payment type.

  - used for 'Property' and 'Own Damage' payments

- **GB2 regression**:

$$f_P(p) = \frac{|\alpha| p^{\alpha\gamma_1-1}\beta^{\alpha\gamma_2}}{B(\gamma_1,\gamma_2)(\beta^{\alpha}+p^{\alpha})^{\gamma_1+\gamma_2}},$$

with $\alpha \neq 0$, β, γ_1, $\gamma_2 > 0$, $B(\alpha_1,\alpha_2)$ the usual beta function and $\beta_{ijk} = \exp\left(\boldsymbol{x}_{ij}^{'}\boldsymbol{\beta}_{P,k}\right)$.

  - used for 'Injury' payments

Positive part of payments

- Inspired by the histograms of the positive payments, we used a **mixture of lognormal regression** models:

$$\log(P) \sim w_1 N_1(\mu_1, \sigma_1^2) + w_2 N_2(\mu_2, \sigma_2^2) + w_3 N_3(\mu_3, \sigma_3^2),$$

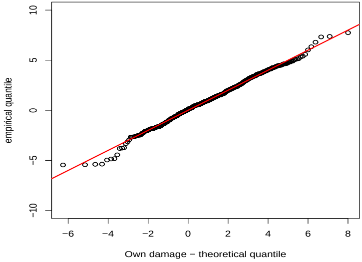
  where w_1, w_2 and w_3 are weights, specified as

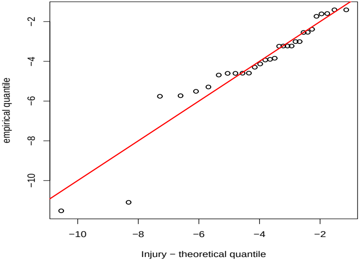
$$\begin{aligned} w_1 &= \frac{\exp(a)}{\exp(a) + \exp(b) + \exp(c)}, \\ w_2 &= \frac{\exp(b)}{\exp(a) + \exp(b) + \exp(c)}, \\ w_3 &= \frac{\exp(c)}{\exp(a) + \exp(b) + \exp(c)}, \end{aligned}$$

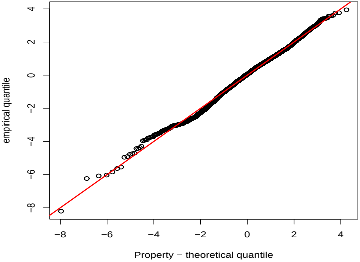
  and $N_i(\mu_i, \sigma_i^2)$ is a normal distribution with mean μ_i and variance σ_i^2.

- Covariate information is incorporated in the weights and parameters μ_i and σ_i^2 ($i = 1, 2, 3$).

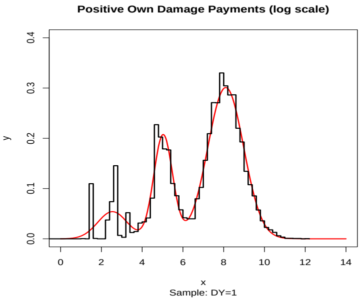
QQ plots on the negative payments

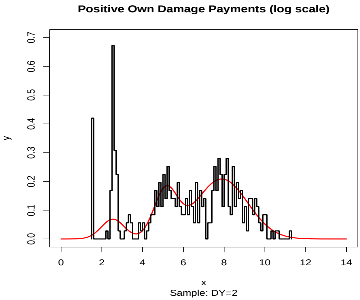


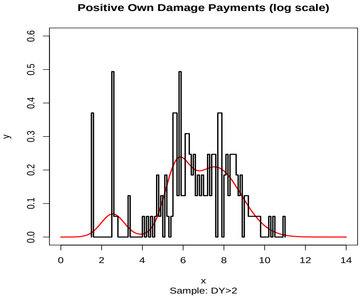




Histograms of the positive payments - own damage



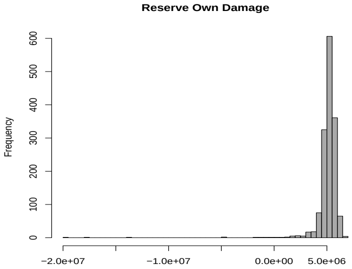


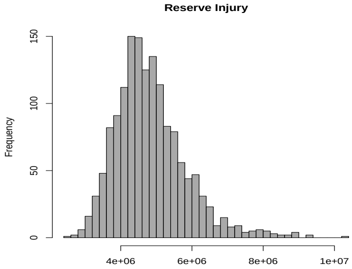


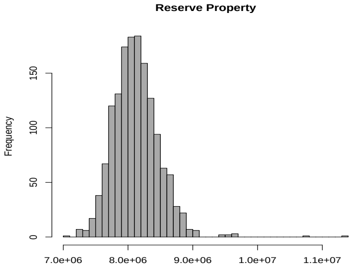
Prediction of RBNS claim reserves

- **Step 1**: *simulate the next event's time interval*
- **Step 2**: *simulate the exact time of the next event*
- **Step 3**: *simulate the event type*
- **Step 4**: *simulate payment type*
- **Step 5**: *simulate payments*
- **Step 6**: *stop or continue, if necessary - depending on whether settled or not*

Resulting predictive distributions of reserves - by type







Concluding remarks

- Main idea: *claims reserving using statistics for recurrent events*.
- The hope is to improve the prediction of reserves using detailed micro-level recorded information.
    - the cost is the additional complexity in the modeling involved.
- Additional work to be done:
    - comparing the results with traditional reserving methods.
- Similar methodology to other areas of actuarial statistics e.g. recurrent episodes in workers' compensation.

Thank you!