Reserving Methods Using Individual Claim-Level Data

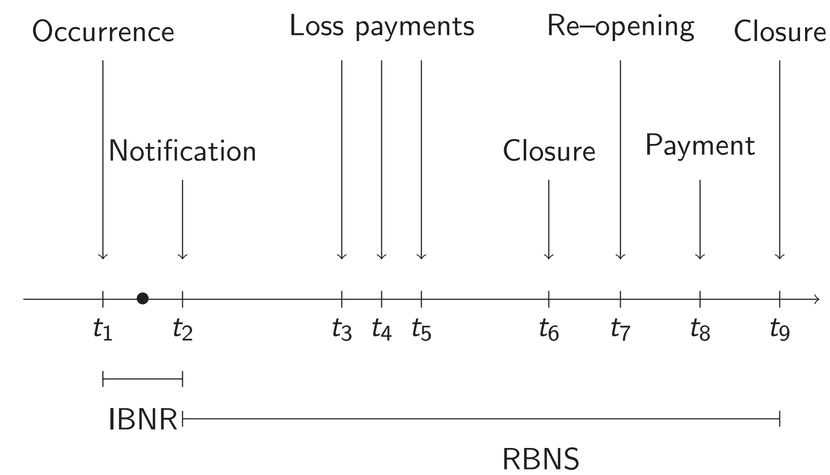
Katrien Antonio

CAS Loss Reserve Seminar
Boston

September 17 2013

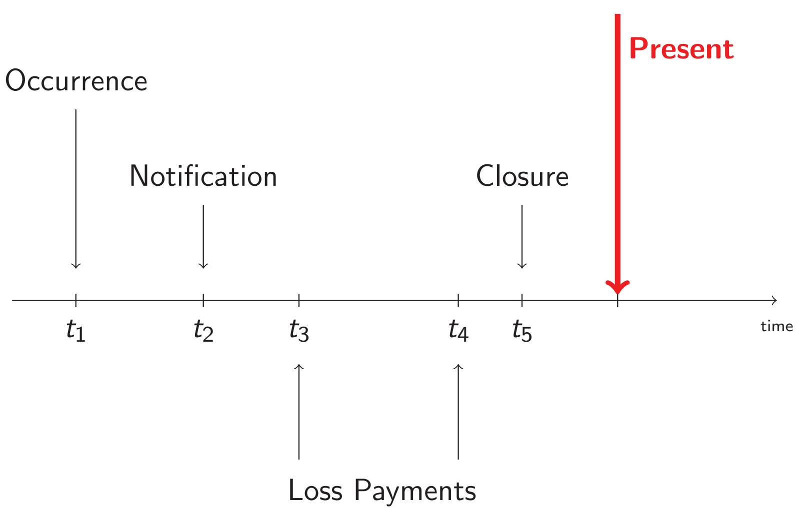
Motivation: dynamics

- **Individual, micro or granular level → claim–by–claim**.



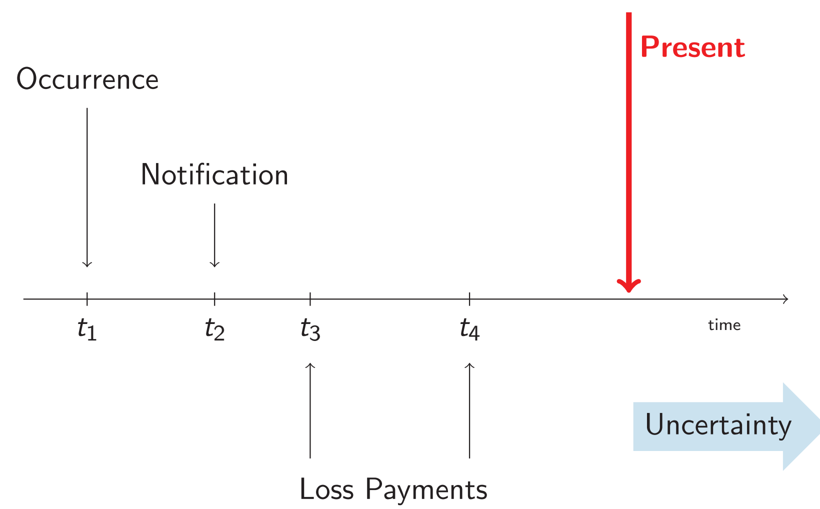
Motivation: dynamics - types of claims

Closed claim = 'closure' (at t_5) $\leq$ present (say, τ).



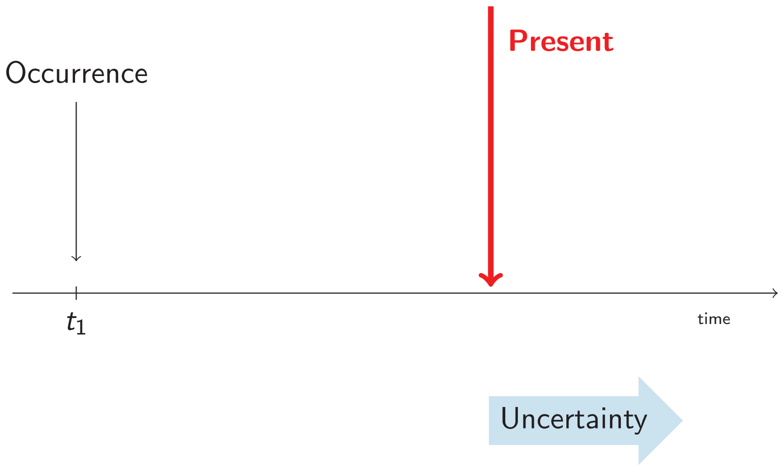
Motivation: dynamics - types of claims

RBNS claim = reported, but 'closure' > present (say, τ).



Motivation: dynamics - types of claims

IBNR claim = incurred, but reporting and ‘closure’ > present (say, τ).



Motivation: from macro to micro–level models

- **Traditional approach**:
  - aggregate data by (arrival, development) year combination;
  - apply reserving method designed for run–off triangle

    (cfr. many of the talks presented at this seminar).
- **Our viewpoint**:

  design a reserving method at individual claim level.
- **Inspiration?**

Motivation: from macro to micro–level models

“The problem is more with the data than the methods, since, clearly, it is **the estimation of aggregate case reserves which is at fault**. [. . .] In this respect, models based on individual claims, rather than data aggregated into triangles, are likely to be of benefit. [. . .] Aggregate triangles are useful for management information, and have the advantage that simple deterministic methods can be used to analyze them.”(England & Verrall, 2002, p.507)

“However, it has to be borne in mind that **traditional techniques** were developed **before the advent of desktop computers**, using methods which could be **evaluated using pencil and paper**.” (England & Verrall, 2002, p.507)

“The triangle is a **summary**, whose origins are very much **driven by the computational restrictions of a bygone era**.” (Taylor & Campbell, 2002, p.21)

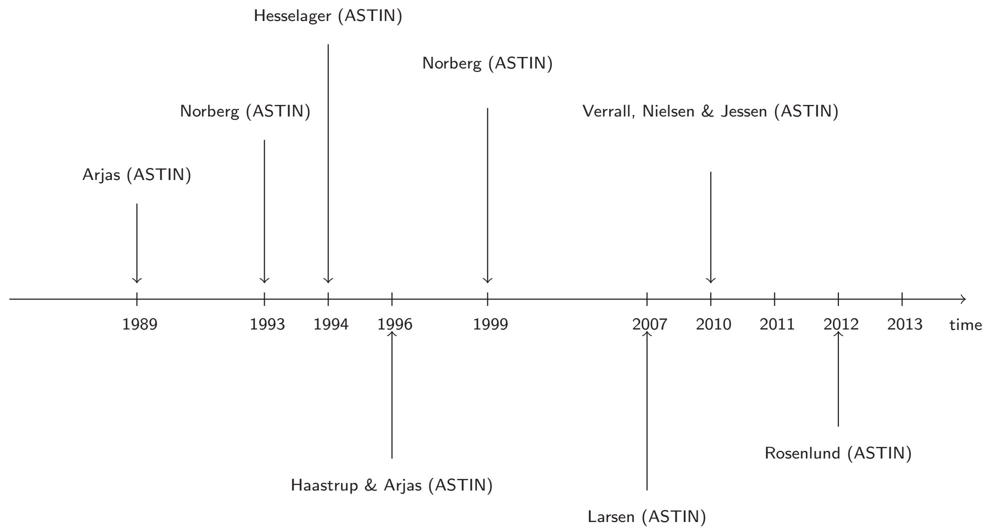
Motivation: from macro to micro–level models

“Recall that condition XXX in Mack’s model had only one purpose: it was chosen in order to explain the form of the chain ladder estimators XXX used by practitioners for predicting future claim numbers. Hence condition XXX was chosen for **pragmatic** reasons.” (Mikosch, 2009, p.374)

“Conditions such as XXX and XXX (i.e. Mack’s conditions) do not explain the dynamics of the underlying claim arrival and payment process in a satisfactory way. This is not surprising since only the first and second conditional moments of the annual dynamics are specified.” (Mikosch, 2009, p.381)

Micro–level reserving: strictly Scandinavian?

(Time line contains a selection of papers from international actuarial journals.)



My research on micro–level loss reserving

- Antonio & Plat [AP] (2013, SAJ, in press): development of individual claims in **continuous time**, **parametric**;

  inspired by Norberg (1993,1999), Haastrup & Arjas (1996), Cook & Lawless (2007);

- Pigeon, Antonio & Denuit [PAD] (2013, ASTIN Bulletin, in press): development of individual claims in **discrete time**, **parametric**;

  inspired by chain–ladder method;

- Antonio & Godecharle: development of individual claims in **discrete time**, historical simulation from empirical data;

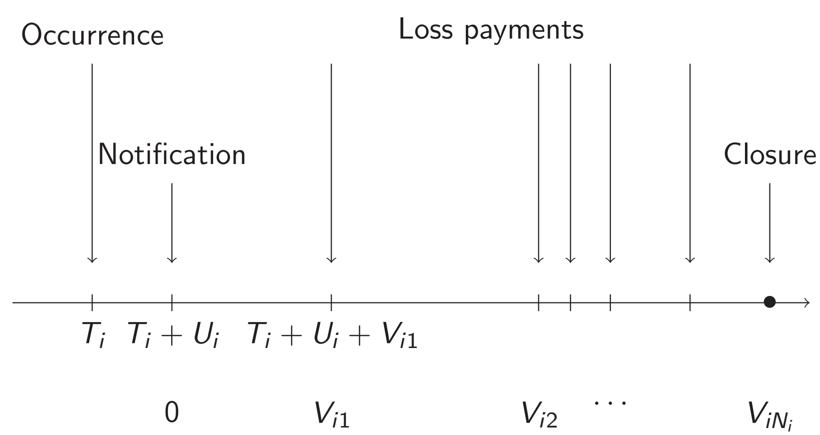
  inspired by Drieskens, Henry, Walhin & Wielandts (Scandinavian Actuarial Journal, 2012) and Rosenlund (ASTIN Bulletin, 2012).

[AP]: continuous time, claim-by-claim

- A **claim** i is a combination of
  - an accident date T_i;
  - a reporting delay U_i;
  - a set of covariates $\mathbf{C}_i$;
  - a development process $\mathbf{X}_i$: $\mathbf{X}_i = (\{E_i(v), P_i(v)\})_{v\in[0,V_{iN_i}]}$;
- In the **development process** we use:
  - $E_i(v_{ij}) := E_{ij}$ the type of the jth event in development of claim i;
  - occurs at time v_{ij}, in months after notification date;
  - corresponding payment vector $P_i(v_{ij}) := P_{ij}$.
- **Event types**?
  - payment, settlement with payment, settlement without payment, ...

[AP]: continuous time, claim-by-claim

Run–off process of a non–life claim: micro–level.



[AP]: statistical model

- Observed data

  development up to time τ of claims reported before τ.

$$(T_i^o, U_i^o, X_i^o)_{i\geq 1}.$$

- Development of claim i is **censored** $\tau - T_i^o - U_i^o$ time units after notification.
- **Likelihood** of the observed claim development process:

$$\propto \left\{\prod_{i\geq 1} \lambda(T_i^o) P_{U|t}(\tau - T_i^o)\right\} \exp\left(-\int_0^\tau w(t)\lambda(t)P_{U|t}(\tau - t)dt\right)$$

$$\cdot \left\{\prod_{i\geq 1} \frac{P_{U|t}(dU_i^o)}{P_{U|t}(\tau - T_i^o)}\right\} \cdot \prod_{i\geq 1} P_{X|t,u}^{\tau - T_i^o - U_i^o}(dX_i^o).$$

[AP]: statistical model - building blocks

- The **reporting delay**: $\prod_{i\geq 1} \frac{P_{U|t}(dU_i^o)}{P_{U|t}(\tau - T_i^o)}$;
- The **occurrence times** (given the reporting delay distribution):

$$\left\{\prod_{i\geq 1} \lambda(T_i^o) P_{U|t}(\tau - T_i^o)\right\} \exp\left(-\int_0^\tau w(t)\lambda(t)P_{U|t}(\tau - t)dt\right);$$

- The **development process** – **event** part:

$$\prod_{i\geq 1} \prod_{j=1}^{N_i} \left\{h_{se}^{\delta_{ij1}}(V_{ij}) \cdot h_{sep}^{\delta_{ij2}}(V_{ij}) \cdot h_p^{\delta_{ij3}}(V_{ij})\right\} \exp\left(-\int_0^{\tau_i}(h_{se}(u) + h_{sep}(u) + h_p(u))du\right);$$

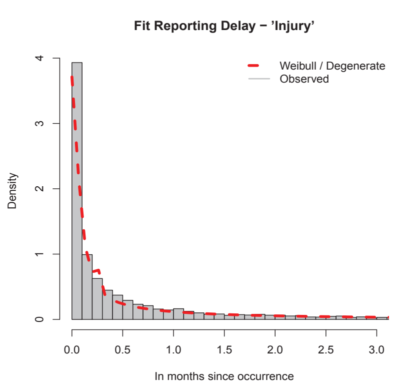
- The development process – **severity** part:

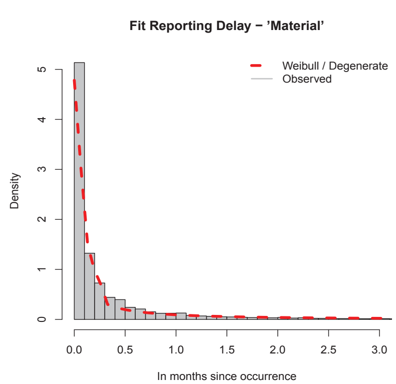
$$\prod_{i\geq 1} \prod_j P_p(dV_{ij}).$$

[AP]: building block - reporting delay

- **Reporting delay**: Weibull with degenerate components at 0 days delay, 1 day delay, ..., 8 days delay:

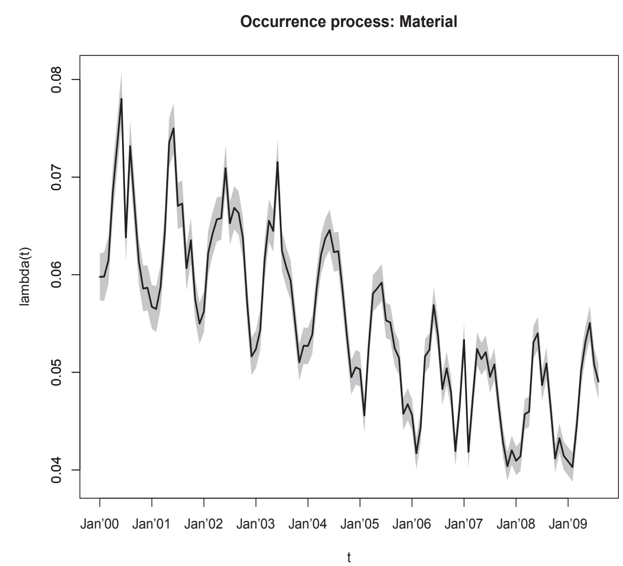
$$\sum_{k=0}^{8} p_k I_{U=k} + (1 - \sum_k p_k) f_{U|U>8}(u).$$



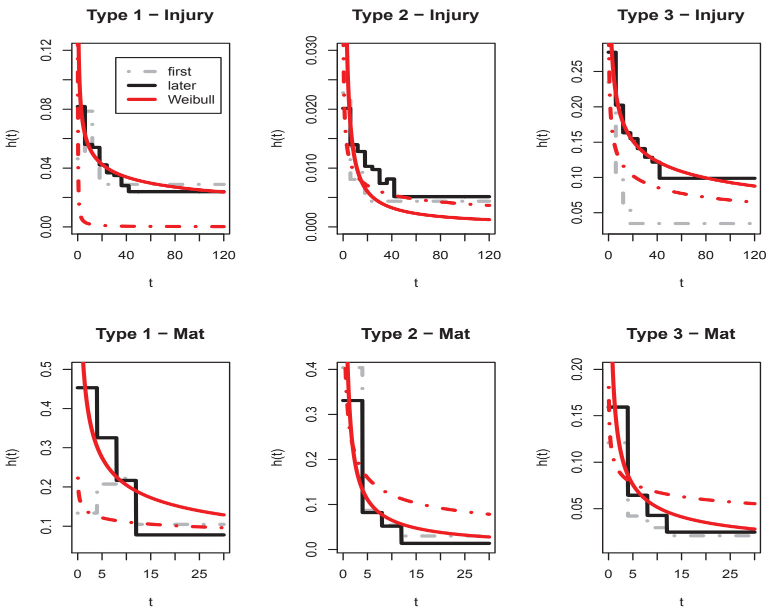


[AP]: building block - occurrence of claims

- Poisson process driving the **occurrence** of Material Damage (MD) claims.



[AP]: building block - occurrence, type of events



[AP]: building block - severity model

- **Severity** distribution.
- Lognormal distributions with μ and σ depending on:
  - the development period: 0-12 months after notification, 12-24 months ... (for injury) and 0-4 months, 4-8 months ... (for material);
  - the initial reserve (set by company experts): categorized.
- Policy limit of 2,500,000 euro is implemented.

[AP]: simulating the predictive distribution of reserves

- Using these building blocks we can easily:
  - **simulate** the time to a next event, the corresponding type and severity for an RBNS claim;
    - use survival function/cdf determined by estimated hazard rates;
  - **simulate** the number of IBNR claims that will show up, their occurrence time and their development;
    - use filtered Poisson process, in combination with reporting delay distribution;

[AP]: example with data from practice

- Data from Antonio & Plat (2013, SAJ):
  - portfolio of general liability insurance policies for private individuals;
  - use data from 1997 to 2004 as training set, 2005-2009 as validation set;
  - exposure measure available;
  - all payments discounted to 1/1/1997;
  - **Bodily Injury** (BI) and **Material Damage** (MD) payments;
  - 279,094 reported claims: 273,977 are MD, 5,117 are BI;
  - closed?: 268,484 MD and 4,098 BI claims.

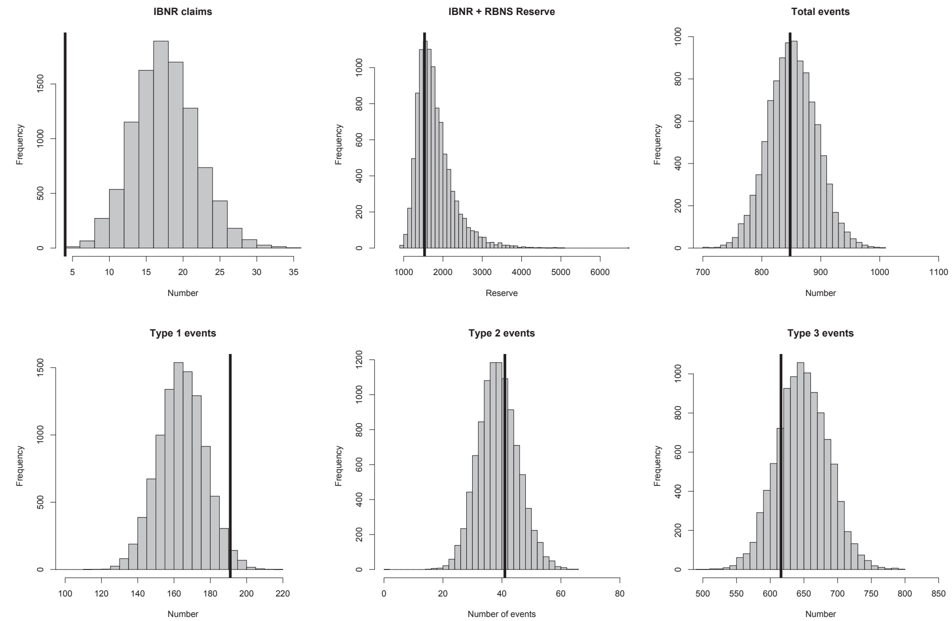
[AP]: example with data from practice

Perform a back–test for BI and MD claims.

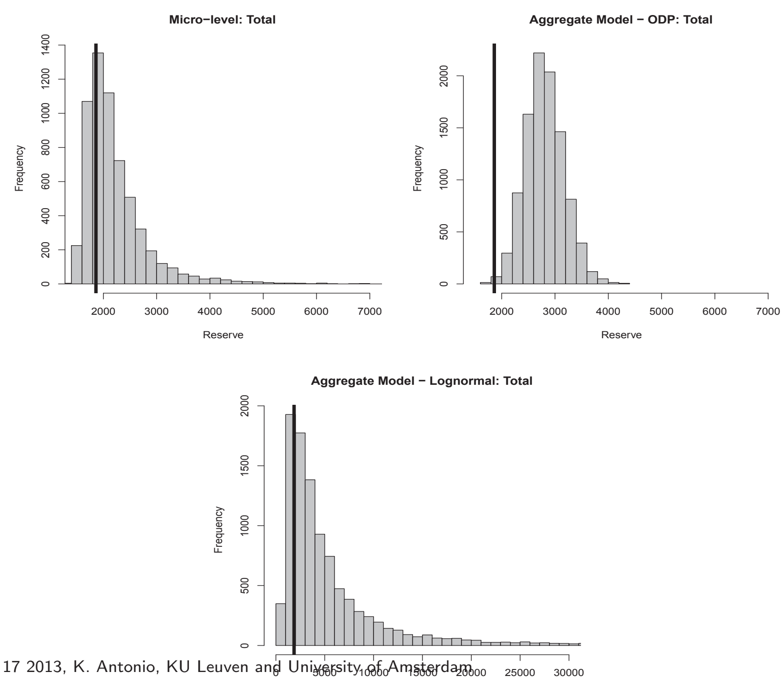
Arrival year	Development year 1	2	3	4	5	6	7	8
1997	308	635	366	530	549	137	132	339
1998	257	482	312	336	269	56	179	**78**
1999	292	590	410	273	254	286	**132**	**97**
2000	317	601	439	498	407	**371**	**247**	**275**
2001	466	846	566	567	**446**	**375**	**147**	**240**
2002	314	615	540	**449**	**133**	**131**	**332**	**1,082**
2003	304	802	**617**	**268**	**223**	**216**	**173**	
2004	333	**864**	**412**	**245**	**273**	**100**		

Total outstanding BI = 7,923,000 and total outstanding MD = 1,861,000.

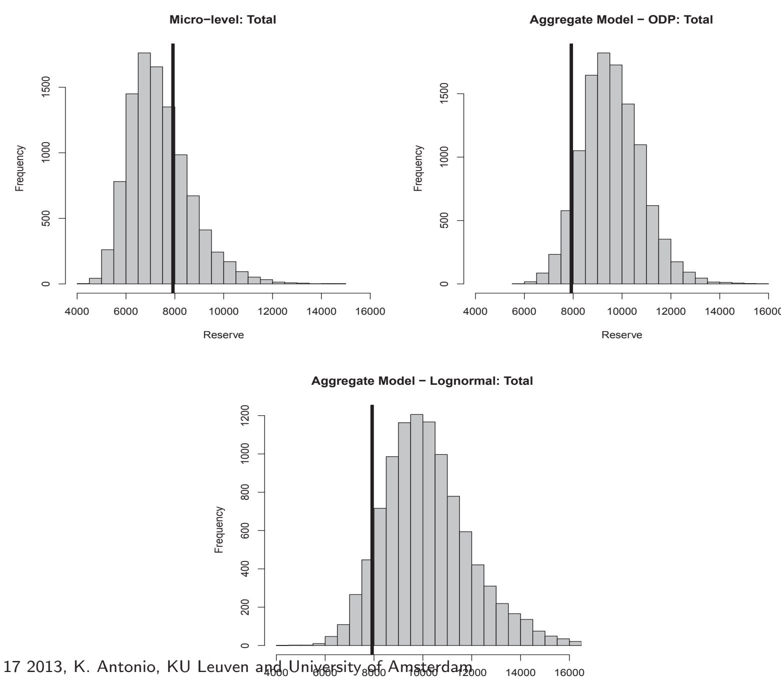
[AP]: output for BI claims, CY 2006



[AP]: MD claims, collective vs. micro–model



[AP]: BI claims, collective vs. micro–model



From continuous time to discrete time, claim–by–claim

- Viewpoint in Antonio & Plat (2013, SAJ): continuous time, inspired by survival analysis.
- Pigeon, Antonio & Denuit (2013, ASTIN) switch to discrete time, inspired by chain–ladder.

[PAD]: discrete–time, an example

	Date	Our notation	
Accident	06/17/1997	$i =$ '1997'	
Reporting	07/22/1997	$t_{(ik)} = 0$	
Number of periods with payment > 0 after first one		$u_{(ik)} = 4$	
Payment	09/24/1997	$q_{(ik)} = 0$	
	10/21/1997		
	11/07/1997		$Y_{(ik),1}$
	05/08/1998	$n_{(ik),1} = 1$	
	12/11/1998		$Y_{(ik),2}$
	03/23/1999	$n_{(ik),2} = 1$	$Y_{(ik),3}$
	02/23/2000	$n_{(ik),3} = 1$	$Y_{(ik),4}$
	01/03/2001	$n_{(ik),4} = 1$	
	02/24/2001		$Y_{(ik),5}$
Closure	08/13/2001	$n_{(ik),5} = 0$	

[PAD]: model set–up

We identify (in discrete time):

- (ik) claim k from occurrence period i;
- T_{ik} **reporting delay**, i.e. number of periods between occurrence and notification period;
- Q_{ik} **first payment delay**, i.e. number of periods between notification and first payment period;
- U_{ik} number of periods with partial payment > 0 after first payment;
- Y_{ikj} the jth incremental partial amount for claim (ik) (> 0);
- N_{ikj} delay between 2 periods with payment, i.e. number of periods between payments j and $j+1$;
- $N_{ik,U_{ik}+1}$ number of periods between last payment and settlement.

[PAD]: development pattern

- We use:
  - initial amount Y_{ik1};
  - payment–to–payment development factors (**note: chain–ladder is period–to–period**)

$$\lambda_j^{(ik)} = \frac{\sum_{r=1}^{j+1} Y_{ikr}}{\sum_{r=1}^{j} Y_{ikr}}, \quad j = 1, \ldots, u_{ik}.$$

- Thus, development pattern $\mathbf{\Lambda}_{u_{ik}+1}^{(ik)}$ is

$$\mathbf{\Lambda}_{u_{ik}+1}^{(ik)} = \begin{bmatrix} Y_{ik1} & \lambda_1^{(ik)} & \ldots & \lambda_{u_{ik}}^{(ik)} \end{bmatrix}'.$$

- **Note**: multivariate distribution of $\mathbf{\Lambda}_{u_{ik}+1}^{(ik)}$ should account for dependence and asymmetry $\Rightarrow$ use Multivariate (Skew) Normal.

[PAD]: likelihood expressions

- As in [AP] we can write down likelihood contributions of:
  - **occurrence** of claims: use Poisson process, thinned;
  - **closed** claims;
  - **RBNS** claims: u_{ik} is not observed, development is censored;
  - **RBNP** claims (**new!**): Reported But Not Paid
    - first period with payment not observed (yet);
    - first payment delay is censored, etc.
  - **IBNR** claims: use Poisson process, appropriately thinned;
- We use:
  - (not MSN) maximum likelihood;
  - (in MSN) maximum product of spacings for shape, combined with ML for location and scale parameters.

[PAD]: what we get – analytical results

- An **IBNR or RBNP** claim: $C = Y_1 \cdot \lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_U$

  with

$$E[\text{IBNR}|\mathcal{I}] \text{ versus } E[\text{RBNP}|\mathcal{I}]$$
$$= (x) \cdot E_U \left[2^{U+1} \exp(\mathbf{t}_1' \boldsymbol{\mu}_{U+1} + 0.5 \mathbf{t}_1' \boldsymbol{\Sigma}_{U+1}^{1/2} \left(\boldsymbol{\Sigma}_{U+1}^{1/2} \right)' \mathbf{t}_1) \cdot \prod_{j=1}^{U+1} \Phi \left(\frac{\Delta_j \cdot ((\boldsymbol{\Sigma}_{U+1}^{1/2})' \mathbf{t}_1)_j}{\sqrt{1+\Delta_j^2}} \right) \right],$$

  where (x) is $E[K_{\text{IBNR}}]$ versus k_{RBNP}.

- An **RBNS** claim: $[C|\boldsymbol{\Lambda}_A = \boldsymbol{\ell}_A] = y_1 \cdot \ell_1 \cdot \ldots \cdot \ell_{u_A - 1} \cdot \lambda_{u_A} \ldots \cdot \lambda_U$

  with

$$E[\text{RBNS}|\mathcal{I}]$$
$$= \sum_{(ik)_{\text{RBNS}}} y_1 \cdot \ell_1 \cdot \ldots \cdot \ell_{u_1 - 1}$$
$$\cdot \quad E_{U_B} \left[2^{U_B} \exp(\mathbf{h}_1' \boldsymbol{\mu}_{U+1}^* + 0.5 \mathbf{h}_1' \left(\boldsymbol{\Sigma}_{U+1}^* \right)^{1/2} \left(\left(\boldsymbol{\Sigma}_{U+1}^* \right)^{1/2} \right)' \mathbf{h}_1) \cdot \prod_{j=1}^{U_B} \Phi \left(\frac{\Delta_j^* \cdot \left(\left(\left(\boldsymbol{\Sigma}_{U+1}^* \right)^{1/2} \right)' \mathbf{h}_1 \right)_j}{\sqrt{1 + \left(\Delta_j^* \right)^2}} \right) \right].$$

[PAD]: what we get – predictive distributions

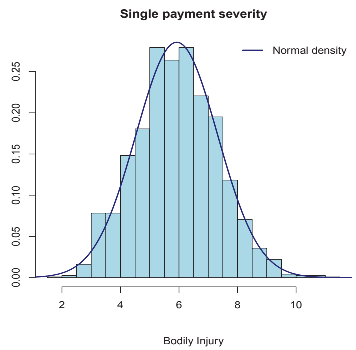
- Simulation–based, using all building blocks;
- Parameter uncertainty?
  - simulate each parameter from asymptotic normal distribution;
  - (at least for discrete time components and location parameter in M(S)N);
  - more work to be done.

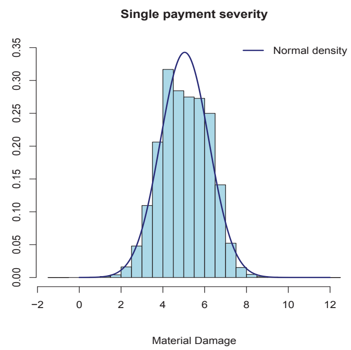
[PAD]: example with (same) data from practice

- Distribution for **number of periods**: $\{T_{ik}, Q_{ik}, U_{ik}, N_{ik}\}$
  - mixtures of discrete distribution with degenerate components;
  - model selection using AIC and BIC;
  - comparison of observed data and fits.
- **Development pattern**:
  - Multivariate Normal (MN) and Multivariate Skew Normal (MSN) with UN, TOEP, CS and DIA structure for $\boldsymbol{\Sigma}_c^{1/2}$;
  - model selection with AIC and BIC;
  - comparison of observed data and fits.

[PAD]: example with (same) data from practice

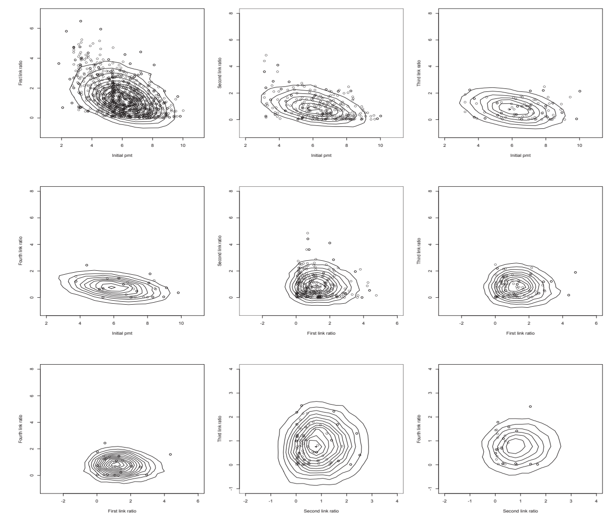
- Claims closing with 'single' payment: **explore U(S)N**.





[PAD]: example with (same) data from practice

- Claims with multiple payments: **explore M(S)N**.



[PAD]: example – predictive results

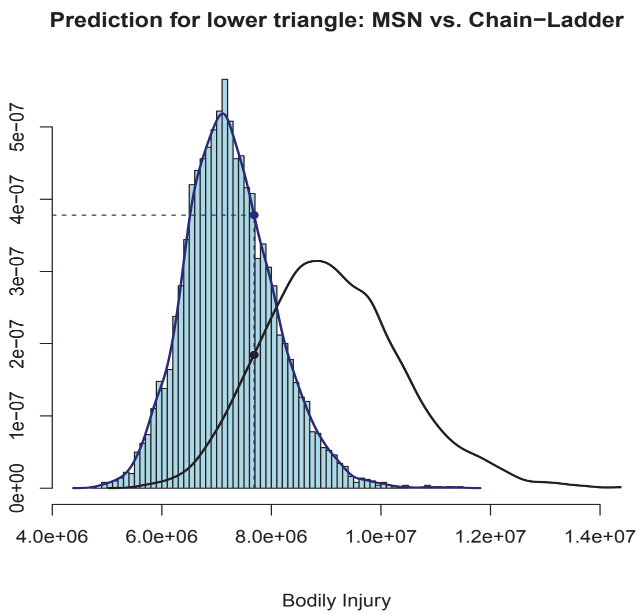
Model or Scenario	Item	Expected Value	S.E.	$\text{VaR}_{0.95}$	$\text{VaR}_{0.995}$
Individual MSN	IBNR^+	2,970,645			
Analytical	RBNS	5,433,548			
(until settlement)	Total	8,404,192			
Individual MSN	IBNR^+	3,035,519	494,771	3,912,159	4,673,340
Simulated	RBNS	5,439,318	704,701	6,650,958	7,738,003
(until settlement)	Total	8,474,837	853,812	9,927,439	11,105,174
Chain-Ladder (Bootstrap, ODP)	Total	9,126,639	1,284,793	11,380,743	13,061,937
Observed	Total	7,684,000			

([PAD] uses slightly adjusted distinction MD vs BI, compared to [AP].)

[PAD]: example – predictive results

Model or Scenario	Item	Expected Value	S.E.	$\text{VaR}_{0.95}$	$\text{VaR}_{0.995}$
Individual MSN Sim. + Unc. (until settlement)	Total	8,568,506	922,657	10,134,198	11,406,905
Individual MSN Sim. + Unc. + Pol. Limit (until settlement)	Total	8,568,355	902,601	10,141,226	11,320,931
Individual MSN Sim. + Unc. + Pol. Limit (until triangle bound)	Total	7,251,103	817,878	8,679,618	9,717,771
Chain-Ladder (Bootstrap, ODP)	Total	9,126,639	1,284,793	11,380,743	13,061,937
Observed	Total	7,684,000			

[PAD]: example – predictive results



Highlights

- Novel setting for claims reserving at individual level.
- Continuous time framework, inspired by survival analysis, on the one hand.
- Discrete time framework, inspired by chain–ladder, on the other hand.
- Analytical expressions for moments of (RBNS, RBNP, IBNR) reserve + simulation approach.
- More work ongoing.
- Comments?: k.antonio@uva.nl.