Actuaries Institute

Correlations between insurance lines of business: An illusion or a real phenomenon?

Benjamin Avanzi, School of Risk and Actuarial Studies, UNSW and Département de Mathématiques et de Statistique, Université de Montréal

Greg Taylor, Bernard Wong, School of Risk and Actuarial Studies, UNSW

Insights, 23 October 2015

Actuaries Institute

INTRODUCTION

The authors and their Linkage Project

- Authors from School of Risk and Actuarial Studies, UNSW
- They hold a Linkage Grant awarded by the Australian Research Council
  - Subject: "Modelling claim dependencies for the general insurance industry with economic capital in view…"
  - Term: 3 years+
  - Collaborative between, and jointly funded by Government, industry (Allianz, IAG, Suncorp) and academia
- This presentation relates to one of the many projects funded by the Grant
- Based on a paper available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2597405

PROLOGUE

Dependency between lines of business (LoBs)

- Relevant to **diversification**, as it affects:
  - Risk margins
  - Capital margins
- Risk margins
  - V@R 75%: centre of distribution: **(Pearson) correlation** a reasonable measure of dependency
  - V@R 99.5%: right tail of distribution: correlation unlikely to be helpful, some measure of **tail dependency** more useful
  - This presentation concerned with correlation and risk margins

Cross-LoB correlations: “conventional wisdom”

- Published papers on numerical values of cross-LoB correlations are:
  - Bateup & Reed (2001)
  - Collings & White (2001)
- Some insurers may rely on other proprietary work, but the above papers form, in some sense, an industry benchmark

Cross-LoB correlations: "conventional wisdom" – example 1

- Bateup & Reed: **total correlation** for OSC

	ABI	Workers Comp	Prof Ind	Inwards Re	Fire/ ISR	APD	Home	Other
Liab	0.25	0.25	0.25	0.25	0	0	0	0
ABI		0.35	0.25	0.25	0	0.25	0	0
Workers Comp			0.25	0.25	0	0	0	0
Prof Ind				0.25	0	0	0	0
Inwards Re					0.05	0.05	0.05	0.05
Fire/ISR						0.10	0.10	0.05
APD							0.20	0.10
Home								0.10

CTP (→ ABI) Motor (→ APD)

Cross-LoB correlations: "conventional wisdom" – example 2

- Bateup & Reed: **correlation** for OSC **systemic variance** only (excludes process error)

	ABI	Workers Comp	Prof Ind	Inwards Re	Fire/ ISR	APD	Home	Other
Liab	0.35	0.40	0.45	0.45	0	0	0	0
ABI		0.50	0.40	0.40	0	0.55	0	0
Workers Comp			0.45	0.50	0	0	0	0
Prof Ind				0.55	0	0	0	0
Inwards Re					0.15	0.15	0.15	0.15
Fire/ISR						0.40	0.35	0.10
APD							0.75	0.25
Home								0.20

Cross-LoB correlations: "conventional wisdom" (continued)

- The example contains some large correlations
  - Many of 0.4 or more
  - Up to a maximum of 0.75
- We do not assert that these correlations are wrong
- Rather that they should be **model dependent**
  - And we consider how changing the model might change the correlations that should be incorporated in these triangles

Layout of presentation

- What should be measured?
- What can be gleaned from theory?
- A simulation example of what can go wrong
- Some examples based on real data
- Modelling the past vs forecasting the future
- Some conclusions

WHAT SHOULD BE MEASURED?

Notation

- Claim array Δ for LoB n
  - Same shape Δ for all n
- Observation in (k, j) cell of Δ is $Y_{kj}^{(n)}$

Pearson correlation (between claim arrays of two LoBs)

- Well known definition

$$r^{(n_1,n_2)} = \frac{T^{-1}\sum_{k,j\in\Delta}\left(Y_{kj}^{(n_1)} - \bar{Y}^{(n_1)}\right)\left(Y_{kj}^{(n_2)} - \bar{Y}^{(n_2)}\right)}{S^{(n_1)}S^{(n_2)}}$$

where

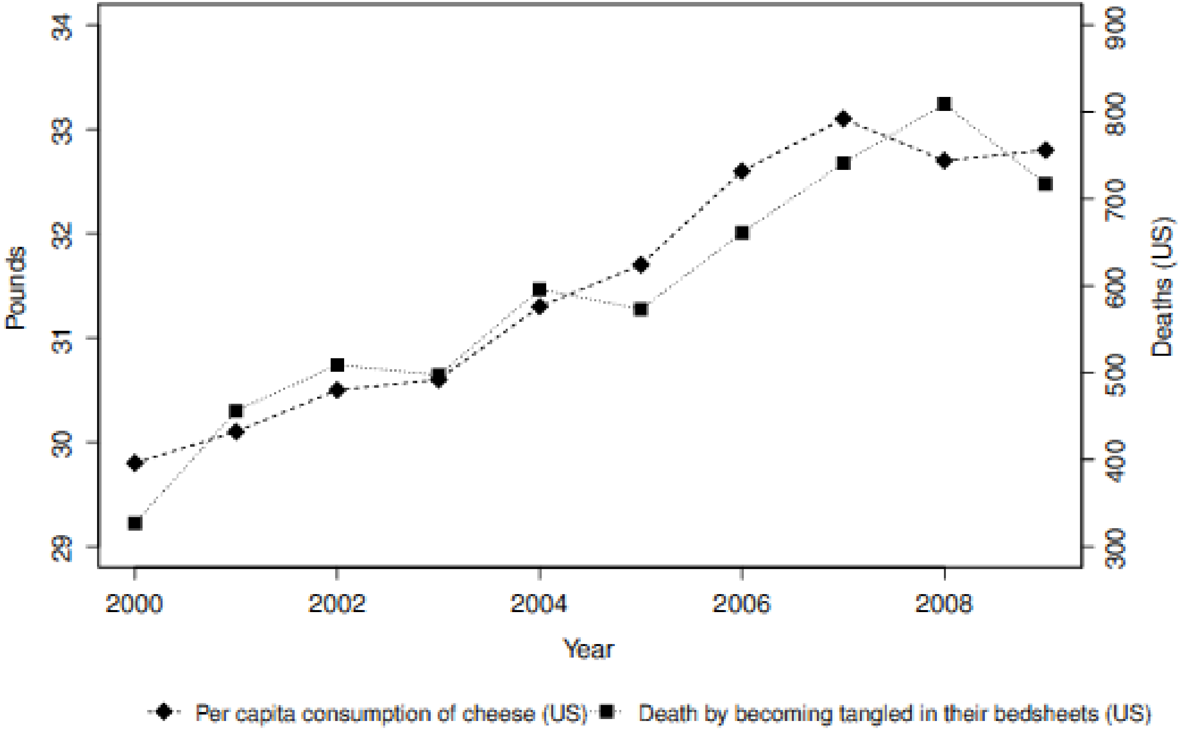
T = number of observations in Δ

$\bar{Y}^{(n)}$ = mean of the observations $Y_{kj}^{(n)}$

$S^{(n)}$ = sample standard deviation of the observations $Y_{kj}^{(n)}$

- **THIS DEFINITION WILL NOT WORK WELL IN OUR CASE WITHOUT MODIFICATION!!!**

Pearson correlation blooper



- Example from http://www.tylervigen.com
- **Correlation** between Per capita consumption of cheese and Deaths by becoming tangled in their bedsheets = **0.95**
- Yet common sense suggests correlation = 0

Pearson correlation blooper (cont'd)

- This calculation would be awarded an F grade in Time Series 101
  - **Rule:** de-trend all time series before calculating correlations
- Why?
  - Otherwise the example tells us only that the trends of the two time series are of similar form (roughly linear)
  - This could have been deduced without any concept of correlation
  - Similar (high) correlations can be obtained from claims (and other financial) data simply because of inflation
- So, correlation calculated **after** de-trending of the time series provides a much more powerful tool
  - Because it measures the sympathy in departures of the two time series from their trends

Back to claim data correlations: how should they be calculated

- In the blooper example
  - Estimating a trend (in this case, perhaps just with respect to time) is equivalent to creating a model
  - Correlations calculated after de-trending are correlations between departures from the models
    - i.e. between residuals
- This is the case for all data sets
  - First, model the data (de-trend) to capture all deterministic effects
  - Calculate some form of residuals (stochastic effects)
  - Correlate the residuals
- Correlation is then a function of stochastic quantities, as it should be

WHAT CAN BE GLEANED FROM THEORY?

Measured correlations are model dependent

- It has been shown that measured correlations are based on residuals
- Residuals are departures from model fitted values
- Residuals are therefore model-dependent
- Correlations are therefore model-dependent

Actuaries Institute

How are measured correlations affected by quality of modelling?

- Let future observations be denoted $Y_{kj}^{*(n)}$ (past $Y_{kj}^{(n)}$)
- Write all the $Y_{kj}^{*(n)}$ as a vector $Y^{*(n)}$
- Prediction error is

$$e^{(n)} = \underbrace{\left[Y^{*(n)} - \mu^{*\mathrm{true}(n)}\right]}_{\text{Process Error}} + \underbrace{\left[\mu^{*\mathrm{mod}(n)} - \hat{\mu}^{*(n)}\right]}_{\text{Parameter Error}} + \underbrace{\left[\mu^{*\mathrm{true}(n)} - \mu^{*\mathrm{mod}(n)}\right]}_{\text{Model Error}}$$

True mean

Model forecast

Model forecast in absence of sampling error

How are measured correlations affected by quality of modelling? (cont'd)

$$e^{(n)} = \underbrace{\left[Y^{*(n)} - \mu^{*\text{true}(n)}\right]}_{\text{Process Error}} + \underbrace{\left[\mu^{*\text{mod}(n)} - \hat{\mu}^{*(n)}\right]}_{\text{Parameter Error}} + \underbrace{\left[\mu^{*\text{true}(n)} - \mu^{*\text{mod}(n)}\right]}_{\text{Model Error}}$$

Model Error:

Small for good models
Large for poor models

- Omission of predictive variables from the model (enlarging model error) shifts some of the signal in the data from measured explanatory effects to perceived random effects (noise)
- If the omitted explanatory variables are common to different LoBs, this is likely to create correlation between the "noise" of those LoBs
- Poor modelling may create apparent correlation where none actually exists
  - And none would be estimated with higher quality modelling

The paper contains an algebraic proof of this result

A SIMULATION EXAMPLE OF WHAT CAN GO WRONG

Actuaries Institute

Simulated data

- Data simulated for 2 LoBs: Home & Motor
- Drawn from following model:

$$
\begin{aligned}
Y_{kj}^{(n)} &\sim \text{Poisson}\left(\mu_{kj}^{(n)}, \phi^{(n)}\right), \qquad n = 1, 2; \\
\mu_{kj}^{(n)} &= \exp\left\{ r_k^{(n)} + s_j^{(n)} + t_{k+j-1}^{(n)} \right\},
\end{aligned}
$$

All observations stochastically independent

Row effect | Column effect | Diagonal effect

- Chain ladder structure with superimposed inflation added

Simulated data (cont'd)

- Quarterly paid loss triangles generated with dimension 41 (same dimension as in later real data sets)
- Mean diagonal effects (superimposed inflation) subject to 3 scenarios:
  - **Scenario 1:** annual rate of 10% in diagonals 17 to 28; other diagonals 3%
  - **Scenario 2:** annual rate of 3% in diagonals 1 to 20; thereafter 10%
  - **Scenario 3:** annual rate of 1% in diagonals 1 to 4; 2% in diagonals 5 to 8; increasing by 1% each 4th diagonal; finally 11% in 41st diagonal
  - Each scenario is common to the Home and Motor LoBs
- 1,000 replicates of each scenario for each LoB
  - So 6,000 triangles in all

Analysis of simulated data

- Reminder of data structure

$$
\begin{aligned}
Y_{kj}^{(n)} &\sim \text{Poisson}\left(\mu_{kj}^{(n)}, \phi^{(n)}\right), \qquad n = 1, 2; \\
\mu_{kj}^{(n)} &= \exp\left\{r_k^{(n)} + s_j^{(n)} + t_{k+j-1}^{(n)}\right\},
\end{aligned}
$$

- Each triangle analysed according to the following model

$$Y_{kj}^{(n)} \sim Poisson\left(\mu_{kj}^{(n)}, \phi^{(n)}\right)$$

$$\mu_{kj}^{(n)} = \exp\left\{r_k^{(n)} + s_j^{(n)}\right\}$$

- It is known that this formulation will produce precisely the same results as the conventional chain ladder
  - Model error introduced: diagonal effects omitted

Analysis of simulated data (cont'd)

- Each of the 6,000 triangles analysed by the above chain ladder model
- Standardized deviance residuals computed for the $\frac{1}{2} \times 41 \times 42$ cells in each triangle
- For each of the 3,000 Home-Motor pairs, Pearson correlation of residuals computed:
  - Over all cells;
  - Separately for each accident quarter (AQ);
  - Separately for each development quarter (DQ);
  - Separately for each calendar quarter (CQ);

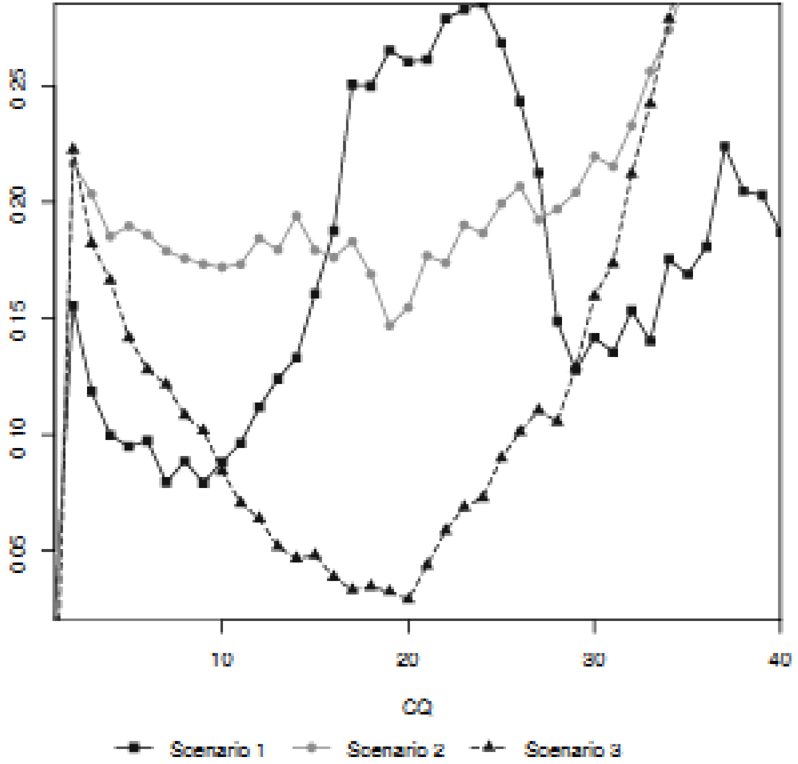
Simulation results (1)

- Home-Motor Pearson correlation across all **cells of triangles**
  - True value = zero
  - Simulated values as follows

Scenario	Pearson correlation
1	+0.20
2	+0.27
3	+0.17

Simulation results (2)

- Simulated Home-Motor correlations by CQ



Superimposed inflation

Scenario 1: High in middle CQs

Scenario 2: High in later CQs

Scenario 3: Steadily increasing over CQs

Simulation results (3)

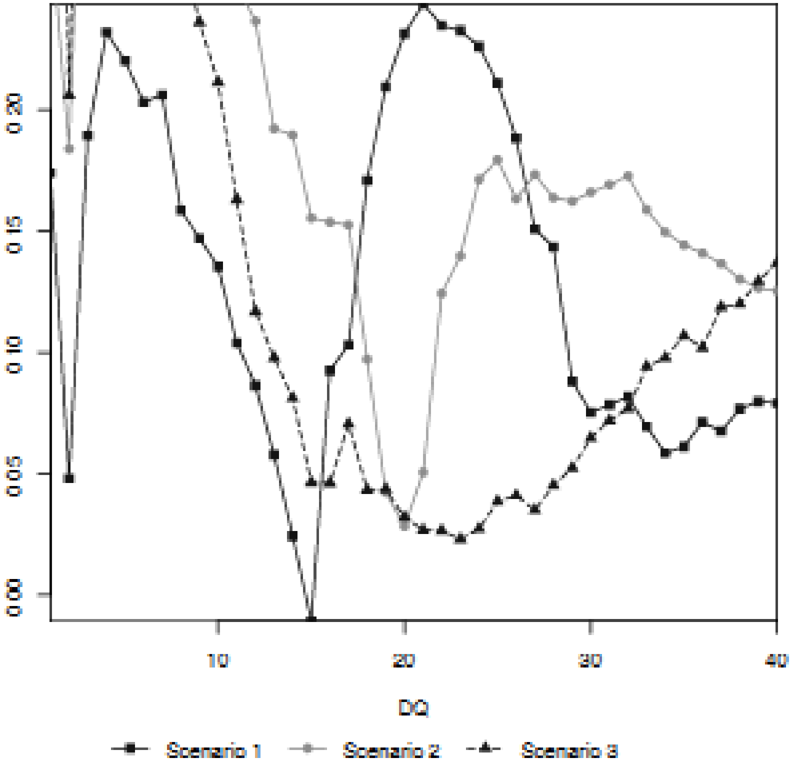
- Simulated Home-Motor correlations by DQ

Superimposed inflation

Scenario 1: High in middle CQs

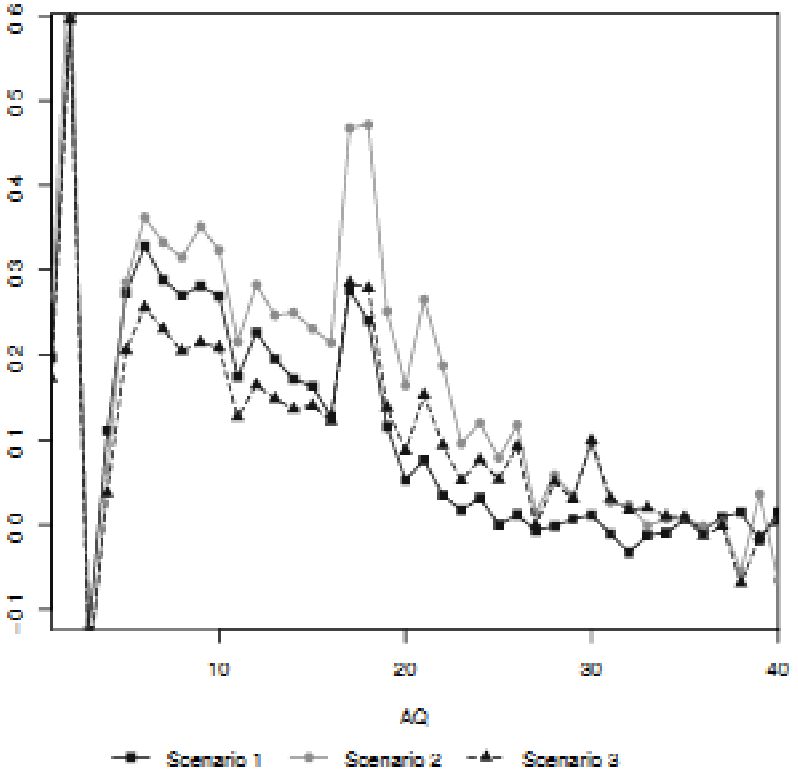
Scenario 2: High in later CQs

Scenario 3: Steadily increasing over CQs



Simulation results (4)

- Simulated Home-Motor correlations by AQ



Superimposed inflation

Scenario 1: High in middle CQs

Scenario 2: High in later CQs

Scenario 3: Steadily increasing over CQs

SOME EXAMPLES BASED ON REAL DATA

Data set

- **AUSI** (**A**llianz, **U**NSW, **S**uncorp, **I**AG)data set
  - Contributed by UNSW's Linkage Project Partners
  - Unit record files for a number of LoBs per Partner
    - Exposure files
    - Claim files
  - Number of years varies by Partner and LoB
    - Up to 10 years for Home and Motor
  - At present 4 LoBs:
    - Home
    - Motor
    - CTP
    - Public Liability

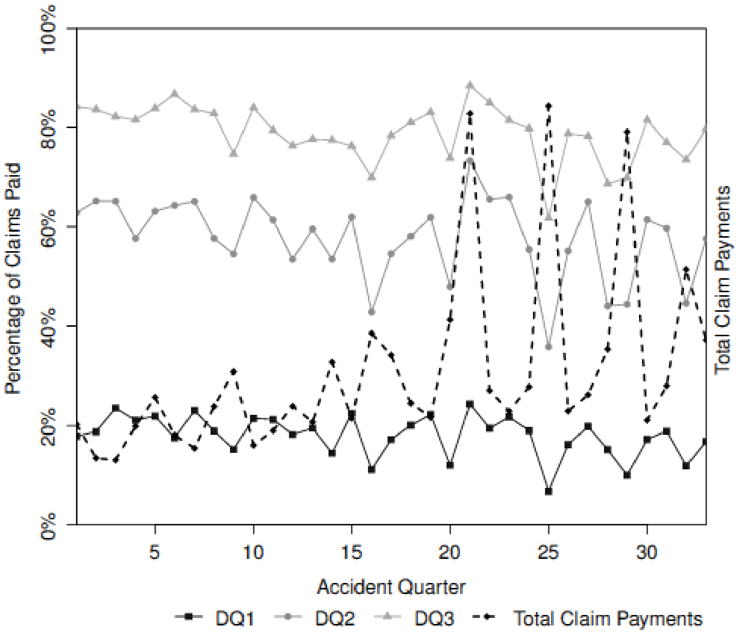
Data analysis

- Each Partner/LoB data summarized in a paid loss triangle
- Each triangle modelled with increasing attention to detail
- For each model
  - Standardized deviance residuals computed
  - Pearson correlations of residuals computed for various LoB pairs within Partner

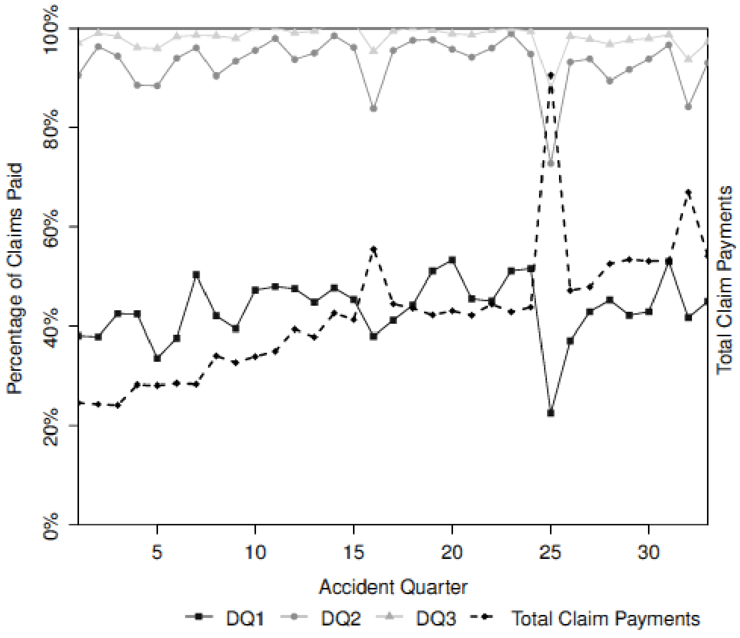
Results of real data analysis (conventional chain ladder)

		Cross-LoB Pearson correlation (whole triangles)					
		Insurer A				Insurer B	
		Home	Motor	CTP	PL	CTP	PL
Insurer A	Home	1	+0.59	+0.04	+0.06		
	Motor		1	+0.04	-0.02		
	CTP			1	-0.02		
	PL				1		
Insurer B	CTP					1	-0.09
	PL						1

Effect of major events?



6.4 (a) Home



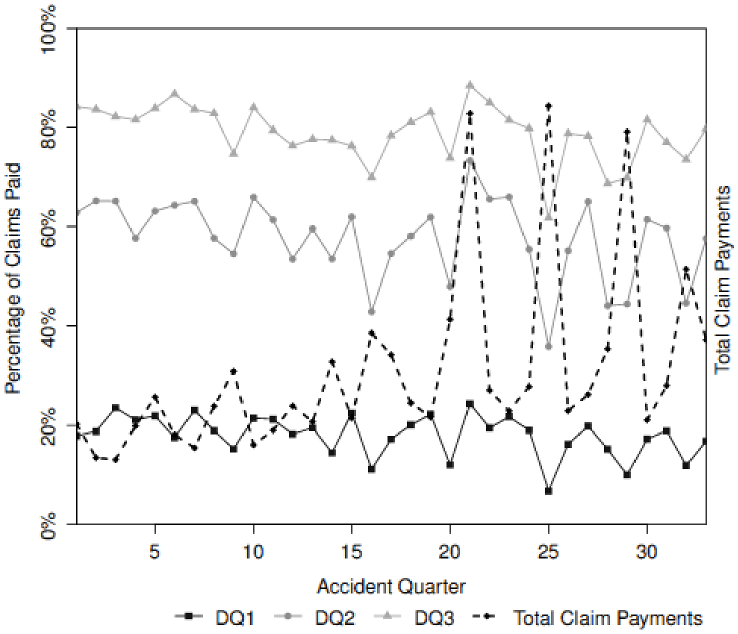
6.4 (b) Motor

- Major events cause sympathetic changes in both LoBs affecting:
  - Volume of claim payments
  - Rate of settle-ment

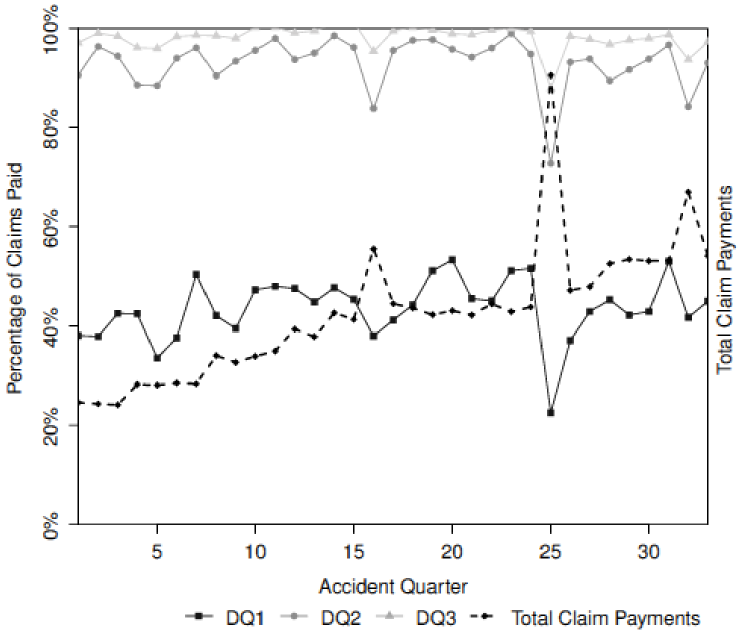
Results of real data analysis (AQs of major events simply deleted from chain ladder)

		Cross-LoB Pearson correlation (whole triangles)					
		Insurer A				Insurer B	
		Home	Motor	CTP	PL	CTP	PL
Insurer A	Home	1	+0.11	+0.04	+0.09		
	Motor		1	+0.02	-0.02		
	CTP			1	-0.02		
	PL				1		
Insurer B	CTP					1	-0.09
	PL						1

Seasonal effects?



6.4 (a) Home



6.4 (b) Motor

- Note seasonal changes in claim volumes
  - Greater in summer (both LoBs)
- Note greater volumes imply slower settlement (both LoBs)

Results of real data analysis (seasonal variates added to chain ladder for DQs 1-3)

		Cross-LoB Pearson correlation (whole triangles)					
		Insurer A				Insurer B	
		Home	Motor	CTP	PL	CTP	PL
Insurer A	Home	1	-0.01	+0.04	+0.09		
	Motor		1	+0.01	-0.02		
	CTP			1	-0.02		
	PL				1		
Insurer B	CTP					1	-0.09
	PL						1

Some observations

- For all pairs of LoBs other than Home-Motor, no statistically significant non-zero correlations are found
  - Even without any attempt to model the esoterica of those LoBs' experience
- Home-Motor requires more care
  - At a superficial level, it exhibits high correlation (0.6)
  - The majority of this is accounted for by a handful of natural events
    - The correlation of experience other than these is low (0.1)
  - This low correlation is accounted for by seasonal factors
    - If the model allows for these, then correlation vanishes

US evidence

- Chain ladder modelling has also been applied to four LoBs in the Meyers-Shi data set that covers many insurers
- Cross-LoB Pearson correlations again computed for 4 LoBs:
  - **PPA:** Private Passenger Auto
  - **CA:** Commercial Auto
  - **WC:** Workers Compensation
  - **OL:** Other Liability

US results

	Pearson correlation (whole triangles)			
	PPA	CA	WC	OL
PPA	1	+0.07	+0.01	+0.06
CA		1	+0.08	+0.00
WC			1	+0.02
OL				1

- Once again, little of interest here
  - Even with crude chain ladder modelling

MODELLING THE PAST VS FORECASTING THE FUTURE

Different types of predictors

- Obervation equation:

$$y = f(predictors) + \varepsilon$$

$$= f(predictors: static, time, unpredictable) + \varepsilon$$

Noise

e.g. AQ

e.g. DQ

e.g. super-imposed inflation

All of these can be modelled in past data (their effects removed)

Certain

Uncertain

Re-inserted into predictions of future cells

NIL

INCREASE: possible cross-LoB correlation

Effect on prediction error

Inferences

- Although it may be possible to model away all cross-LoB correlation in past data
  - It may not be correct to assume zero correlation for the future
  - The extent to which it is incorrect depends on the extent to which unpredictable predictors are included in the model, e.g.
    - Superimposed inflation
    - Major events
    - Claim management changes
    - etc.
  - Again, correlation is model dependent
    - And models of past and future may differ

SOME CONCLUSIONS

Actuaries Institute

Conclusions

1. Cross-LoB dependency is not an absolute
2. It is heavily dependent on the claims models used
3. With some attention to detail, it may be possible to model away virtually all cross-LoB correlation in past data
4. As a very broad generalization:

Better (poorer) modelling → less (greater) perceived dependency

Conclusions (cont'd)

5. A possible (even frequent) consequence of poor modelling is the creation of perceived correlation where none in fact exists
   - This correlation might very well be positive, which would:
     - Reduce measured diversification credit
     - Increase risk margins
     - Increase the insurance risk capital margin
6. Although it may be possible to model away all cross-LoB correlation in past data, it may not be correct to assume zero correlation for the future
   - Consideration will need to be given to allowance for cross-LoB dependency in relation to unpredictable explanatory variables
7. The procedure of modelling away dependency, and then re-inserting part of it
   - Is a more accurate reflection of the real world than failing to model it
   - Will not in general produce the same result as failing to model it

QUESTIONS?