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

Benjamin Avanzi [1,2] Greg Taylor [1] Bernard Wong [1]

[1]School of Risk and Actuarial Studies
UNSW Australia Business School

[2]Département de Mathématiques et de Statistique
Université de Montréal

2017 CAS Annual Meeting – Session C-9*
Anaheim, California, November 6, 2017

*This research was supported under Australian Research Council's Linkage Projects funding scheme (project number LP130100723, with funding partners Allianz Australia Insurance Ltd, Insurance Australia Group Ltd, and Suncorp Metway Ltd). The views expressed here are those of the authors and are not necessarily those of the supporting organisations, nor those of the cosponsors of the program, or those of the employers of the speakers.

Background

Linkage Project 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 UNSW
- This presentation relates to one of the projects funded by the Linkage Project grant.

Paper in May 2016 issue of the ASTIN Bulletin.
Accepted manuscript can be downloaded from UNSWorks.
http://bit.ly/1Ob9Y7B

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

Notation

We focus on claims triangles:

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

Pearson correlation (between claim arrays of two LoBs)

- Well known definition

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

Where

- T = number of observations in Δ
- $\overline{R}^{(n)}$ = mean of the residuals
- S^n = sample standard deviation of the residuals $R_{kj}^{(n)}$

Note: This definition can be improved to allow for different expected volumes in each cell (see paper)

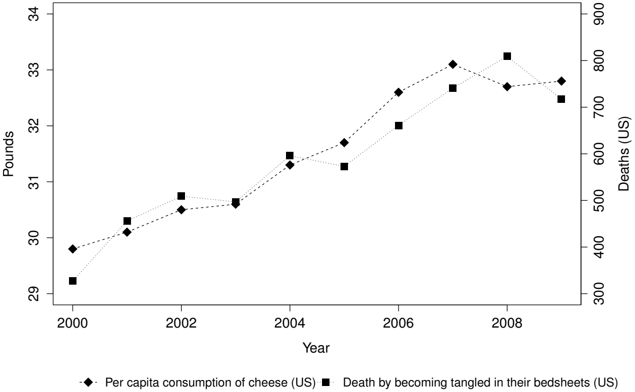
Cross-LoB correlations: “conventional wisdom”

- In the Australian context, 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" (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

Pearson correlation blooper



Correlation = 0.95 !

Example from `http://www.tylervigen.com`

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.

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

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)

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

- 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

Small for good models
Large for poor models

The paper contains an algebraic proof of this result

Data set

AUSI (Allianz, UNSW, Suncorp, IAG) 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

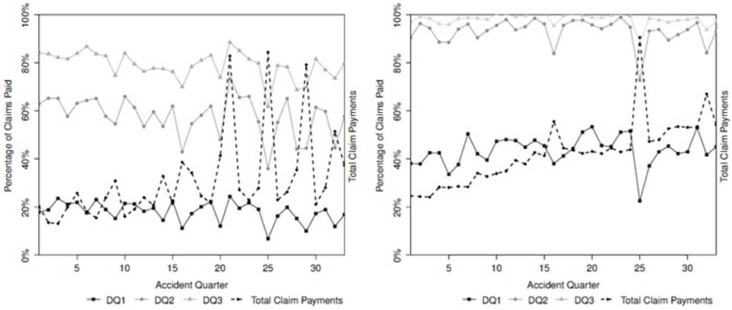
Data analysis

- Each Partner/LoB data summarised in a paid loss triangle
- Each triangle modelled with increasing attention to detail
- For each model
  - Standardised 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

Effects of major events?



6.4 (a) Home   6.4 (b) Motor

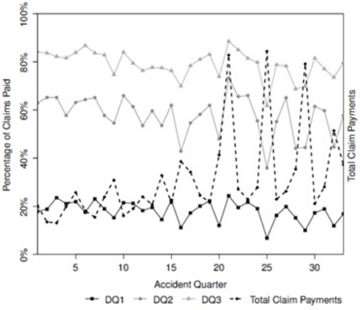
Major events cause sympathetic changes in both LoBs affecting:

- Volume of claim payments
- Rate of settlement

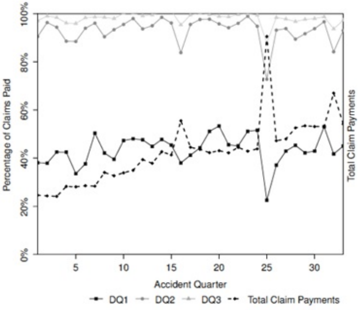
Results of real data analysis (AQs of major events simply deleted from chain ladder)

		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

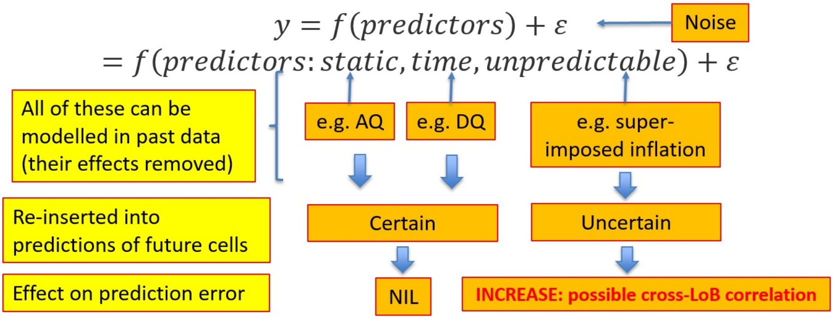
	Pearson correlation (whole triangles)			
	PPA	CA	WC	OL
PPA	1	+0.07	+0.01	+0.06
CA		1	+0.08	+0.00
WC			1	+0.02
OL				1

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

Different types of predictors

Observation equation:

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



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

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
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

Conclusions (cont'd)

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