



Institute  
and Faculty  
of Actuaries

# Model Simplicity

James P. Norman FIA

30 October 2014



Institute  
and Faculty  
of Actuaries

## Bayesian and Information Theoretic Methods for Model Comparison of Copulas, with Applications to Non-Life Internal Models and Reinsurance Pricing

James P. Norman PhD

erlise  
 onorship  
 Thought leadership  
 Progress  
 Community  
 Sessional Meetings  
 Education  
 Working parties  
 Volunteering  
 Research  
 Shaping the future  
 Networking  
 Professional support  
 Enterprise and risk  
 Learned society  
 Opportunity  
 International profile  
 Journals  
 Support

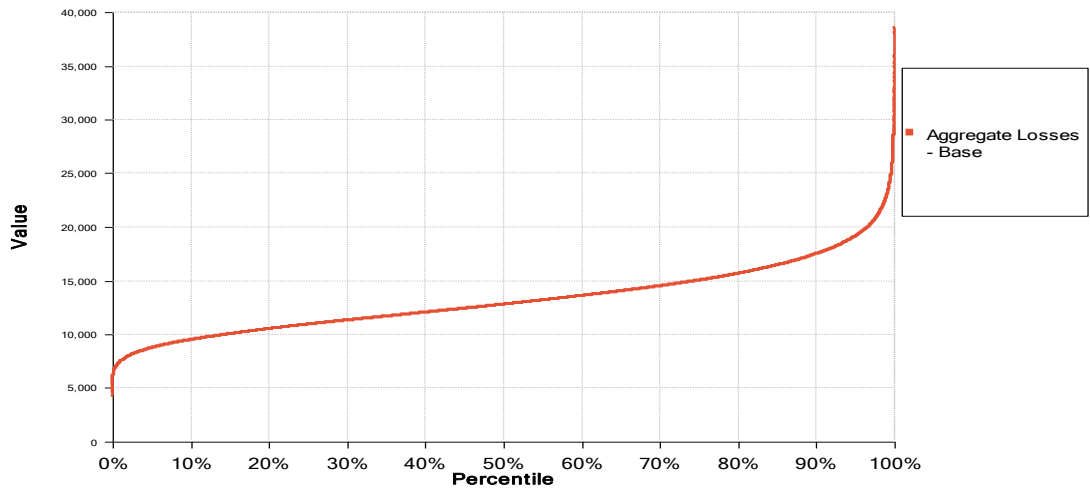
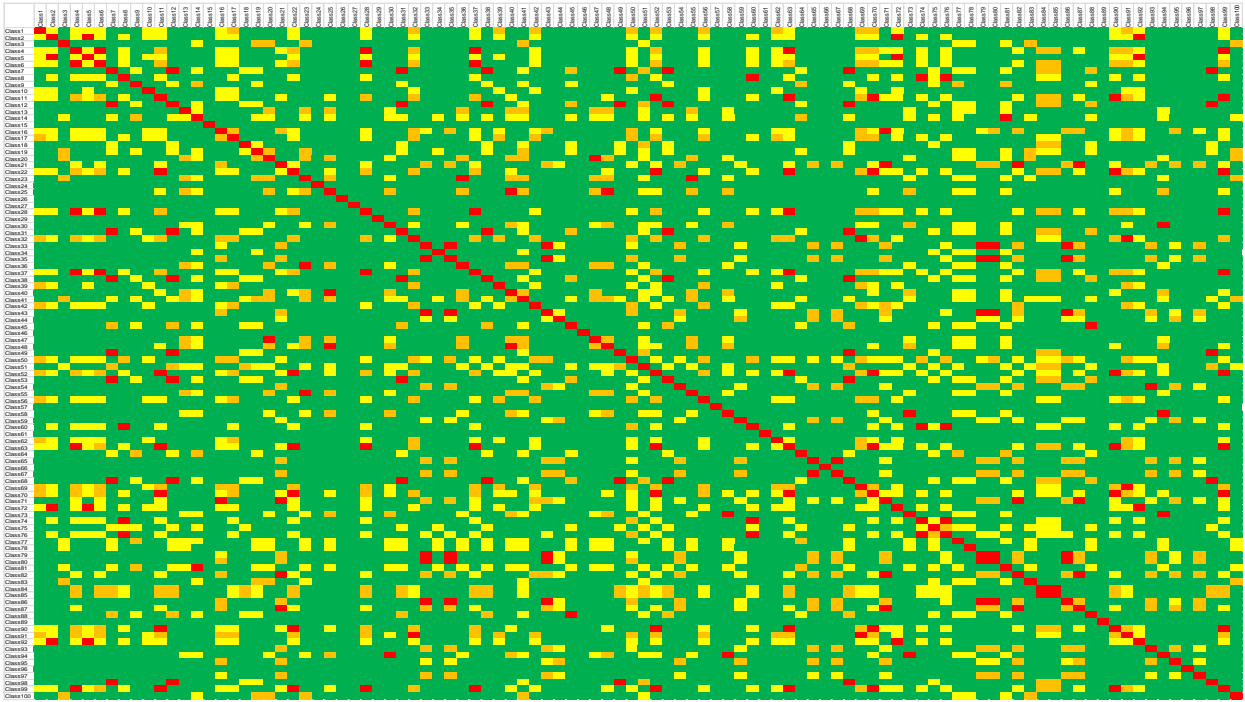
30 October 2014

## Overview

- A “spooky” observation
- The dangers of over-complexity
- Statistical techniques for model selection
  - Akaike
  - Bayes Factors
- Example of choice of copula in small data sets
- Model averaging and Model uncertainty

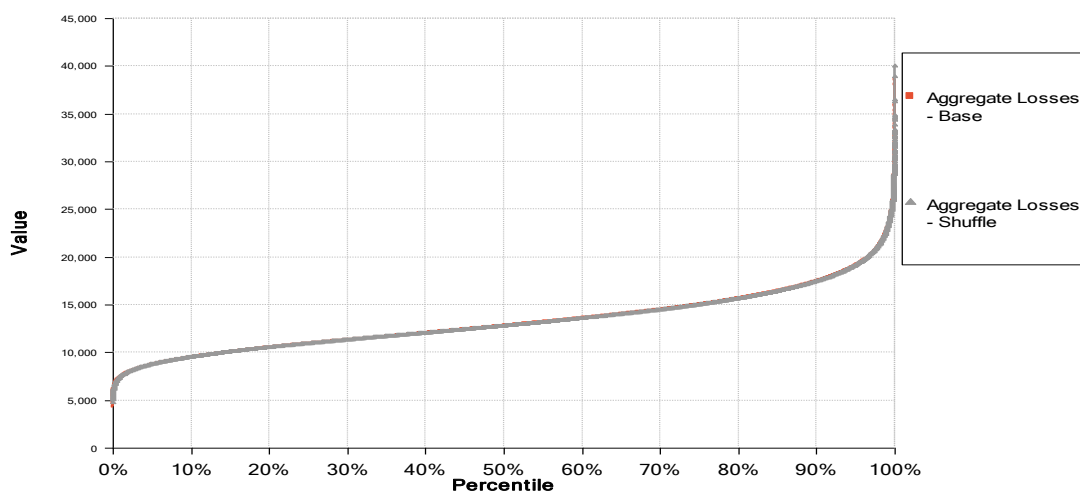
## A “spooky” observation

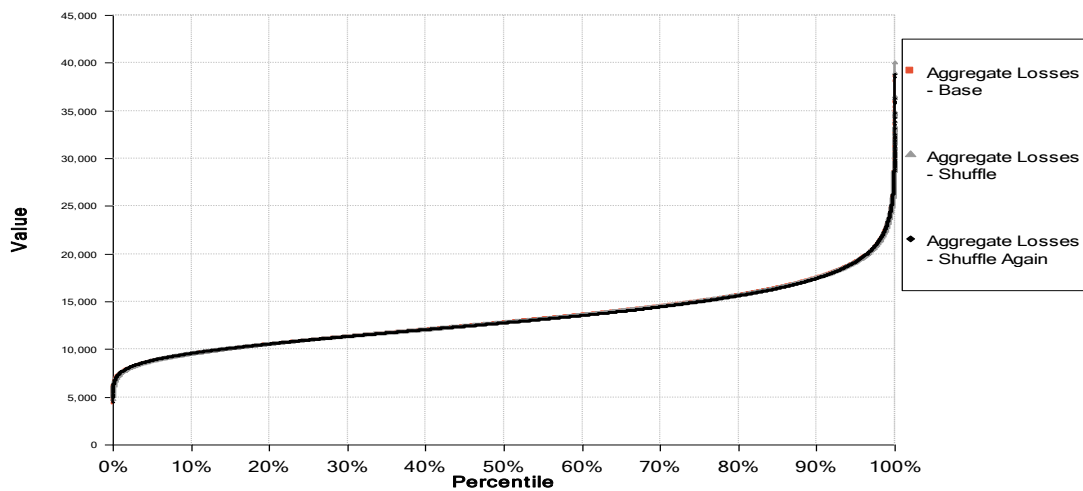
- Example drawn from real P&C internal models
- Gross non-cat claims, 100 lines of business
- Different loss ratio distribution parameters & premium volumes for each line
- Gaussian copula used
  - 4,950 pairwise correlation parameters
- Each parameter estimated through expert judgment, documented and “justified”



## An Experiment

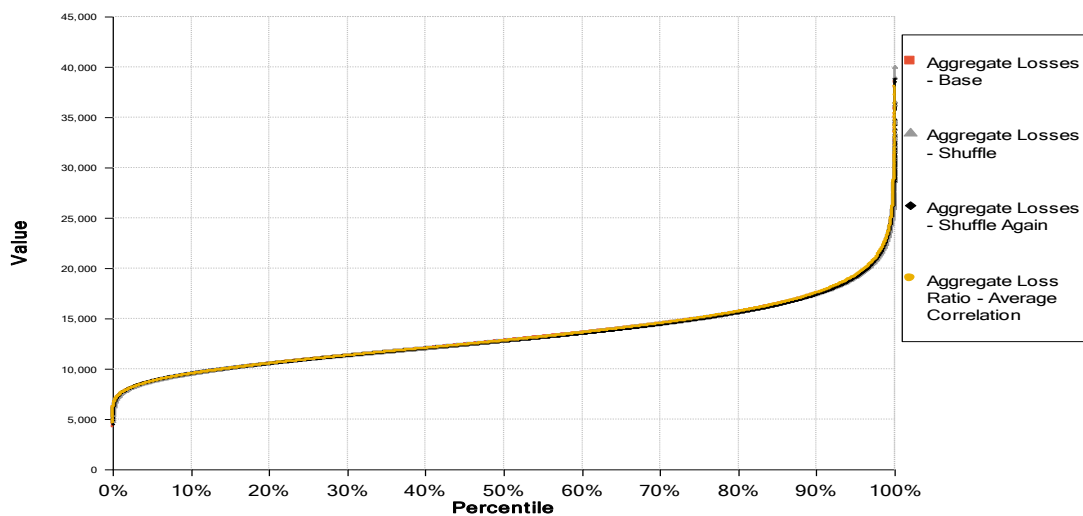
- What do all these parameters do?
- What are the important parameters, for, say, capital?
- An experiment:
  - What if the correlation matrix was scrambled up, so that each pair of lines of business has the “wrong” correlation?
  - Obviously the resulting aggregate claim distribution will be completely wrong...




 | OF Actuaries

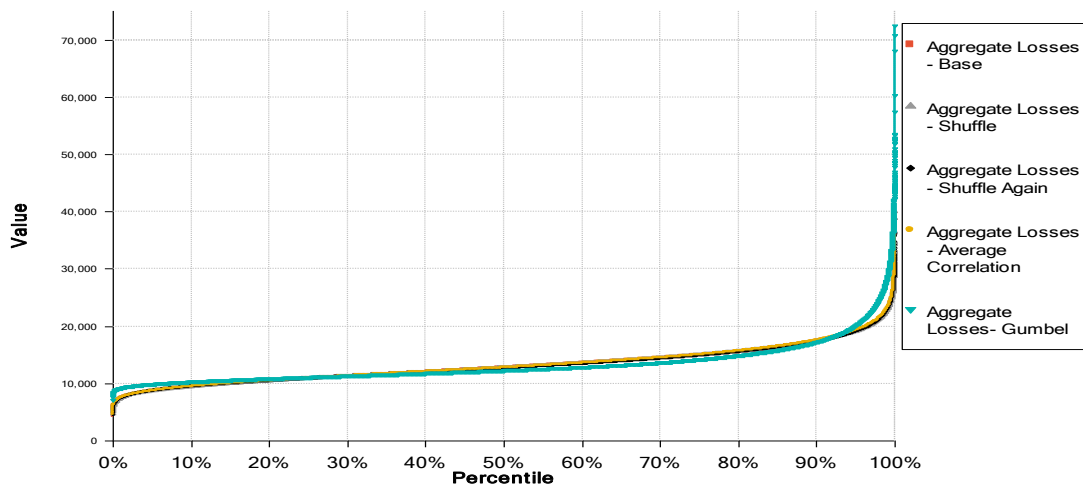
© 2014 Towers Watson. All rights reserved.

towerswatson.com



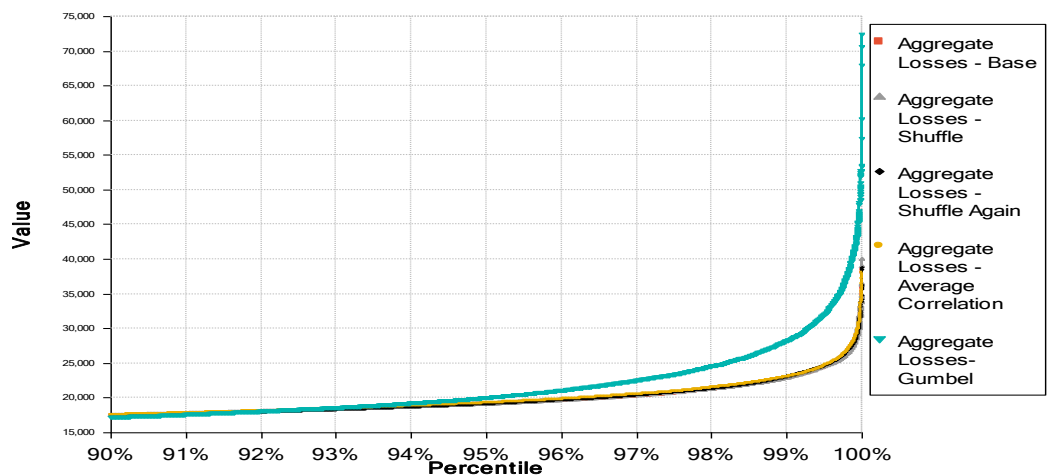
© 2014 Towers Watson. All rights reserved.

towerswatson.com



© 2014 Towers Watson. All rights reserved.

towerswatson.com



© 2014 Towers Watson. All rights reserved.

towerswatson.com

## A “spooky” observation

- The experiments indicate the most important features seem to be:
  - The type of copula
  - The overall “average” level of dependence
- Individual pairwise correlation coefficients seem less important (for the aggregate risk profile)
  - Similar observation in operational risk literature [Brunel, 2014]
- But how much effort goes in to thinking about the first two points?
- Has the model been over-complicated, at the expense of missing the bigger picture?

Brunel V., 2014, “Operational Risk Modelled Analytically”, Risk, July 2014



Institute  
and Faculty  
of Actuaries

30 October 2014

13

## The dangers of over-complexity



Institute  
and Faculty  
of Actuaries

30 October 2014

14

## A quote

*Since all models are wrong the scientist cannot obtain a "correct" one by excessive elaboration. On the contrary following William of Occam he should seek an economical description of natural phenomena. Just as the ability to devise simple but evocative models is the signature of the great scientist so overelaboration and overparameterization is often the mark of mediocrity.*

George E. P. Box



## A more catchy quote

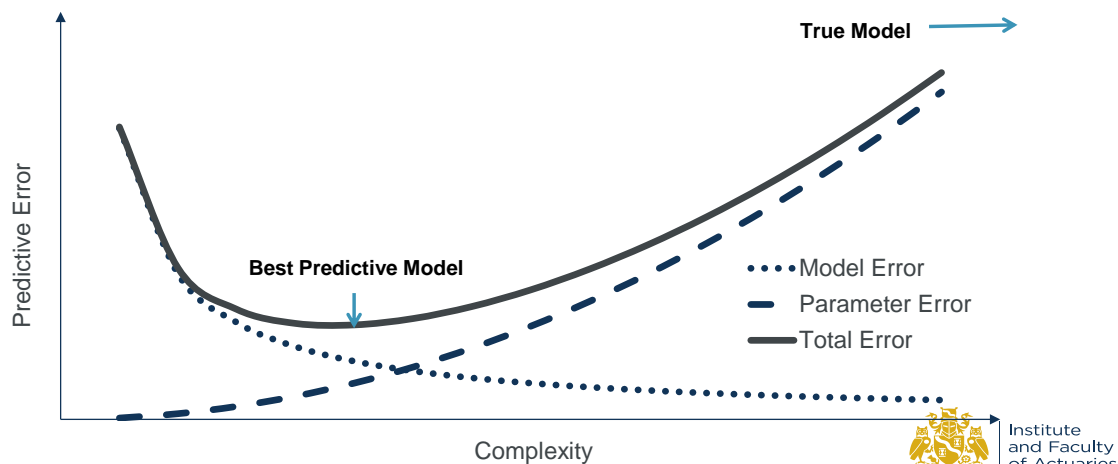
*...all models are wrong, but some are useful.*

George E. P. Box





## The “best” model is not the “true” model



30 October 2014

17

## Statistical Methods for Model Selection

30 October 2014

18

## Statistical Methods for Model Selection

- There are several well known statistical methods for comparing models
- Simplicity/complexity play a key part in most of these
- We will look at two of the most common:
  - Akaike Information Criterion
  - Bayesian Model Comparison



Institute  
and Faculty  
of Actuaries

30 October 2014

19

## Akaike Information Criterion

- Akaike compared the “closeness” of the density of a fitted model  $f_{\hat{\theta}}(y)$  to “truth”  $g(y)$

- “Closeness” measured by Kullback-Leibler information

$$D(f, g) = E_g \left[ \ln \left( g(y) / f_{\hat{\theta}}(y) \right) \right]$$

- An asymptotically unbiased estimator of expected KL information loss is
 
$$-\ln L(\hat{\theta}; \mathbf{x}) + K + \text{const}$$

- K is no. of estimated parameters
- $\ln L(\hat{\theta}; \mathbf{x})$  is maximised log-likelihood of fitted model, given observed data  $\mathbf{x}$
- const is independent of model



Institute  
and Faculty  
of Actuaries

30 October 2014

20

## AIC

$$AIC_i = -2 \ln L(\hat{\theta}; \mathbf{x}, M_i) + 2K_i$$

Model Fit

Complexity

- Model with smallest AIC is expected to have lower predictive error
- Akaike also suggested a weighting scheme to “blend” models:

$$w_i = \frac{e^{-\frac{1}{2}AIC_i}}{\sum_j e^{-\frac{1}{2}AIC_j}}$$

- $w_i$  are “Akaike weights” – interpreted as a *relative weight of evidence* for the model, compared to the other candidate models
- $AIC_c$  is an adjustment to AIC better suited to small data samples



Institute  
and Faculty  
of Actuaries

30 October 2014

21

## Bayesian Model Comparison

- Model structure treated as random with prior distribution updated after data observed
- **Bayes factor** is the ratio of the marginal likelihood of two models

$$B_{12} = \frac{L(\mathbf{x}; M_1)}{L(\mathbf{x}; M_2)}$$

- $B_{12}$  relates prior odds of two models to the posterior odds

$$\frac{p(M_1|\mathbf{x})}{p(M_2|\mathbf{x})} = B_{12} \frac{p(M_1)}{p(M_2)}$$

- $B_{12}$  can be considered the relative weight of evidence in favour of model 1 over model 2 in the data
- By combining the Bayes factor with prior model probabilities, we can obtain posterior probabilities of a model, given the data



Institute  
and Faculty  
of Actuaries

30 October 2014

22

## Marginal Likelihood

- Key quantity is the marginal likelihood of model given data

$$L(M_i; \mathbf{x}) = \int L(\boldsymbol{\theta}; \mathbf{x}, M_i) \pi_i(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

- Can be computed through Monte-Carlo simulation
- Depends on prior  $\pi_i(\boldsymbol{\theta})$  - think carefully
- More complex models generate more complex data, but must spread out their probability mass more widely
- Bayes factors tend to penalise model complexity



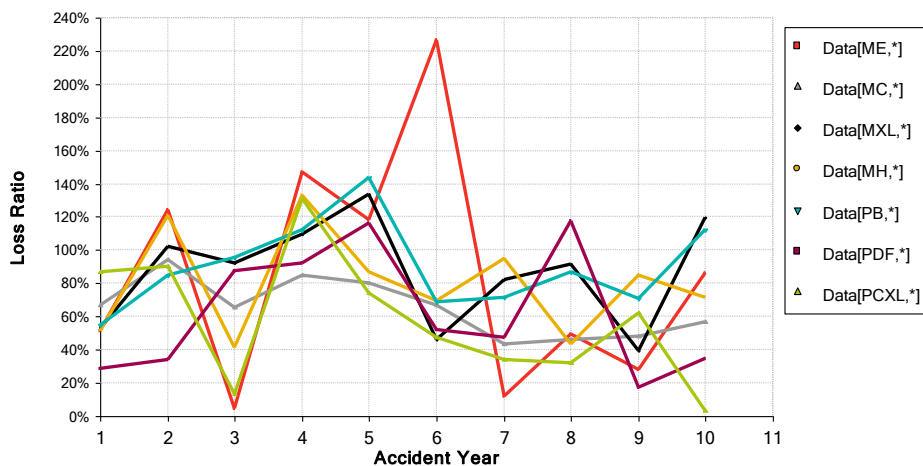
Institute  
and Faculty  
of Actuaries

## Example – which copula?



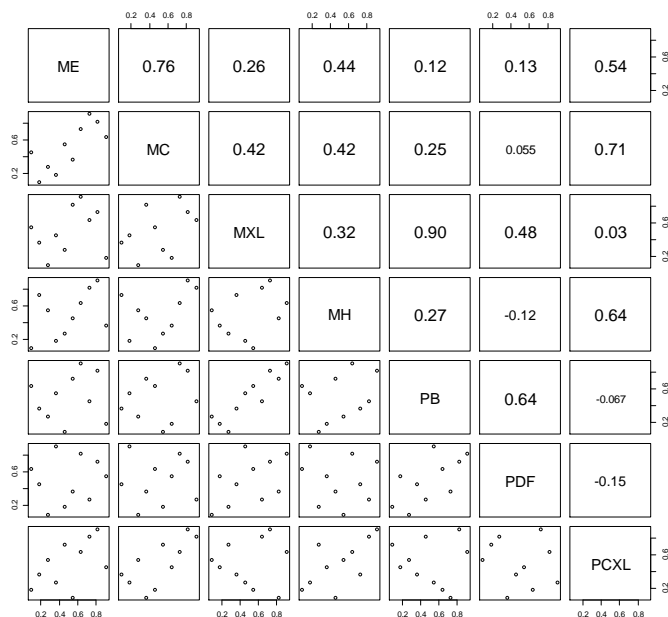
Institute  
and Faculty  
of Actuaries

## Historic Loss Ratio Data



30 October 2014

25

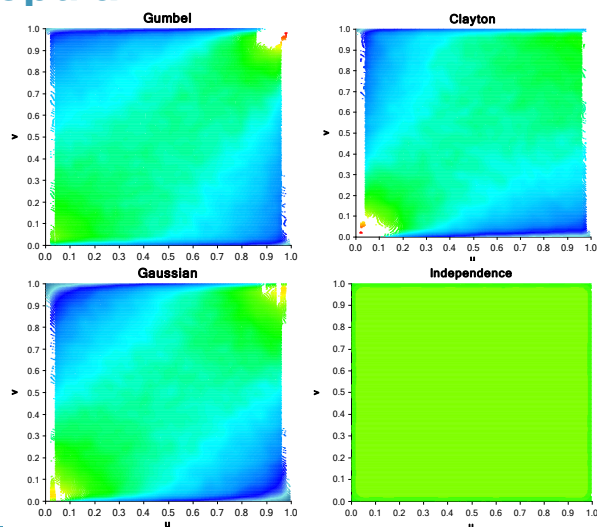


- Rank-Scatter plot matrix and sample Spearman's rank correlation
- Quite a variety of estimated correlations
- Some negative correlations
- Evidence of tail dependence?
  - You gotta be kidding!

30 October 2014

26

## Which Copula?



30 October 2014

27

## Which copula?

- Use gamma distributions for margins
- Compare four models for the copula
  - Gumbel
  - Clayton
  - Gaussian
  - Independence
- Firstly, try pairwise

30 October 2014

28



## AIC Pairwise results

- Independence has highest relative Akaike weight in most pairs
- 2 pairs have highest weight for Gumbel
- Other pairs have relatively even weight between models
- Not really enough data

## Multivariate joint density

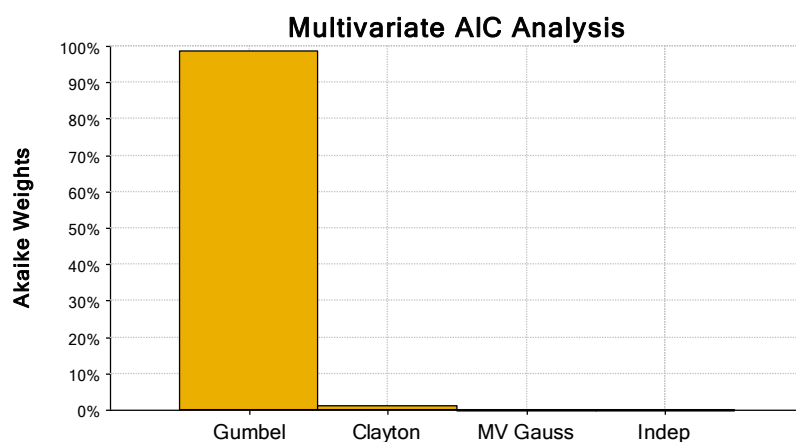
- Rather than model pairwise, consider the joint likelihood of the full multivariate model
- Use the following 7-dimensional multivariate copulas
  - Gumbel (1 parameter)
  - Clayton (1 parameter)
  - Multivariate Gaussian (21 parameters)
  - Independence (0 parameters)
- + 14 parameters for margins



30 October 2014

31

## Akaike Weights for Multivariate Dependence



30 October 2014

32



## Multivariate AIC Results

Copula Model	log-likelihood	K	AIC <sub>c</sub>	$\Delta$ AIC <sub>c</sub>	AIC weights
<b>Gumbel</b>	-6.32	15	57.3	-	<b>98.8%</b>
<b>Clayton</b>	-10.72	15	66.1	8.8	<b>1.2%</b>
<b>MV Gaussian</b>	7.31	35	685.4	628.1	<b>4 x 10<sup>-137</sup></b>
<b>Independence</b>	-14.91	14	70.65	13.34	<b>0.1%</b>

- Gumbel copula is preferred – strong weight indicates substantial support
- Gaussian copula has essentially no support



Institute  
and Faculty  
of Actuaries

30 October 2014

33

## Have we been unfair?

- The Gaussian copula is heavily penalised through having more parameters
  - More parameters  $\rightarrow$  More estimation error  $\rightarrow$  Worse predictive performance
- Gumbel and Clayton have one parameter to model entire multivariate dependence
  - Have 21 times as much data per parameter
- But we want to compare shape as well
- Compare a simpler version of the Gaussian copula, with a single, equal correlation parameter

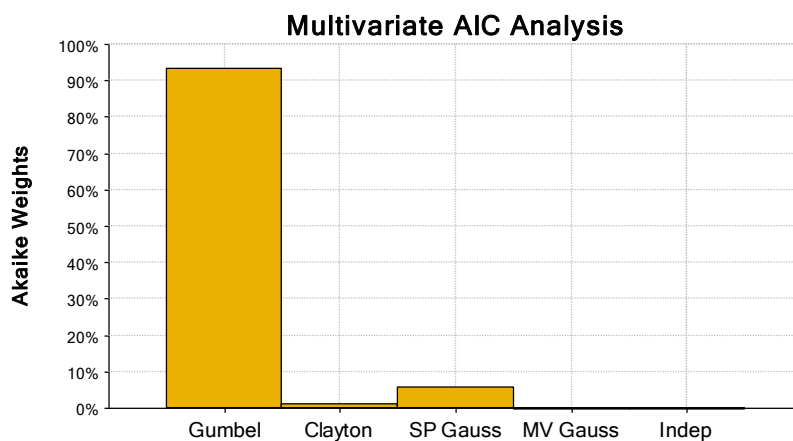


Institute  
and Faculty  
of Actuaries

30 October 2014

34

## With single parameter Gaussian...



## AIC Results

Copula Model	log-likelihood	K	AIC <sub>c</sub>	ΔAIC <sub>c</sub>	AIC weights
Gumbel	-6.32	15	57.32	-	93.2%
Clayton	-10.72	15	66.12	8.80	1.1%
Gaussian – Single Parameter	-9.13	15	62.94	5.62	5.6%
Gaussian – Correlation Matrix	7.31	35	685.37	628.05	4×10 <sup>-137</sup>
Independence	-14.91	14	70.65	13.34	0.1%

## Bayes Factor Calculation

- Use uninformative priors on parameters
  - Uniform  $[0, 1]$  prior on Kendall's  $\tau$  for Gumbel, Clayton and single parameter Gaussian copulas
  - Jointly uniform  $[-1, 1]$  prior on pairwise Kendall's  $\tau$  for Gaussian copula with correlation matrix
    - Restricted to space of PSD correlation matrices
  - Uniform priors on marginal parameters
- Can use informative priors if prior information is available
- Uninformative (equal) prior model probabilities



30 October 2014

37

## Bayesian Model Comparison Results

Copula Model	p(M)	Bayes Factor*	P(M D)
Gumbel	20%	1	82.9%
Clayton	20%	0.022	1.8%
Gaussian - Single Parameter	20%	0.18	14.7%
Gaussian - Correlation Matrix	20%	0.0004	0.04%
Independence	20%	0.006	0.5%

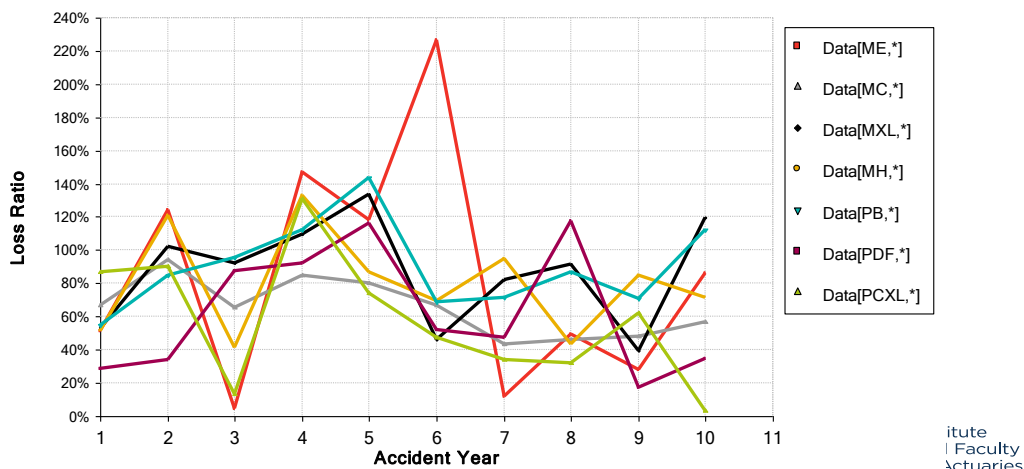
\*Bayes factors relative to the model with Gumbel copula



30 October 2014

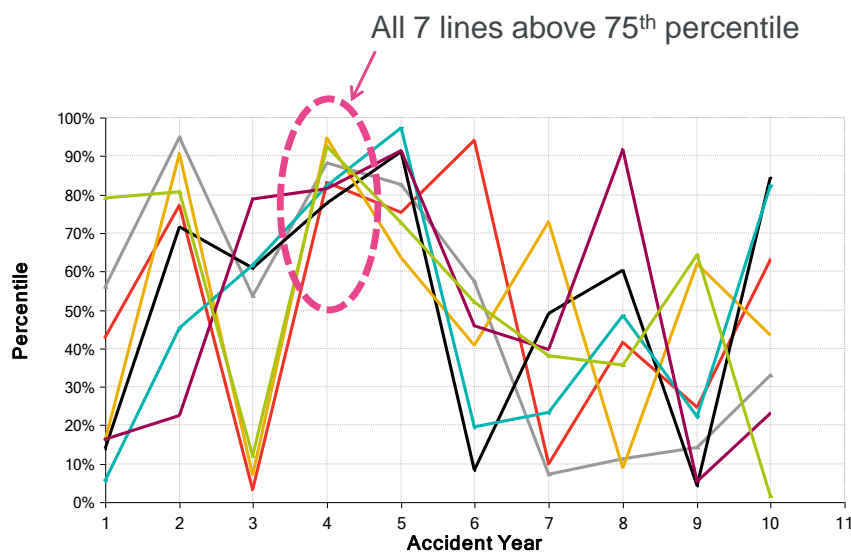
38

## What is driving this?



30 October 2014

39



30 October 2014

40

## Summary Results

- Both AIC and Bayesian results show similar qualitative picture
- Little support for fully parameterised Gaussian copula
- There is, however, information in the data to support and discriminate between **simpler models**
- A reasonably strong weight of evidence for the Gumbel copula, compared to the others
- There could be other models which are better
- **Useful information. To do with as you see fit.**



Institute  
and Faculty  
of Actuaries

30 October 2014

41

## Model Averaging and Model Uncertainty

- “Model uncertainty” – additional uncertainty in forecast distribution due to uncertainty about model structure
- **Model uncertainty can never be completely eliminated**, but can be reduced by blending the results of several models (see e.g. [Bignozzi and Tsanakas 2013])
- “Model averaging” often gives superior predictive results to any of the individual models
- Bayesian approach allows simultaneous quantification of model and parameter uncertainty

Bignozzi, V. and Tsanakas, A. 2013, “*Model Uncertainty in Risk and Capital Measurement*”  
Available at SSRN 2334797



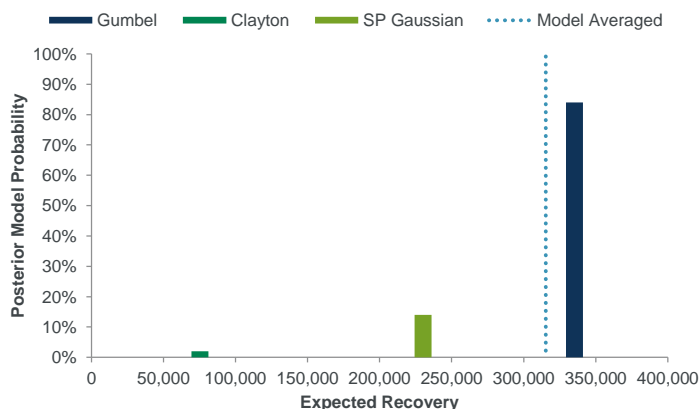
Institute  
and Faculty  
of Actuaries

30 October 2014

42

## Example - Stop Loss

- 25% xs 125% ULR stop loss reinsurance contract



Institute  
and Faculty  
of Actuaries

30 October 2014

43

## Summary and Conclusions

- Include complexity where necessary and justified by evidence
- “Simple” does not mean less sophisticated
- Statistical techniques exist for comparing relative evidence for models
- They can be applied successfully to the problem of copula selection
- There is a surprising amount of information even in small data sets
- Blending models is useful and model uncertainty can be (at least partially) quantified



Institute  
and Faculty  
of Actuaries

30 October 2014

44



# Questions



# Comments

Expressions of individual views by members of the Institute and Faculty of Actuaries and its staff are encouraged.

The views expressed in this presentation are those of the presenter.



Institute  
and Faculty  
of Actuaries