The Actuarial Profession

making financial sense of the future

GIRO 2011

Dimitris Papachristou FSA

Insights into Correlations and Dependencies

- The views expressed in this presentation are my personal views and in particular they should not necessarily be regarded as being those of my employer

Introduction

- Correlation and dependency structures are not intuitive concepts
  - We have some intuition about the likelihood and the effect of a 1% increase in the interest rates on bonds with duration 3 years
  - We do not necessarily have an intuitive understanding of the meaning of a 30% correlation between two lines of business and its effect on the combined capital requirement.
- Purpose of this presentation is to
  - Improve our intuition of correlations and dependencies
  - Investigate the link between common drivers and correlations and dependency structures
  - Look at correlations and dependencies from different angles so that expert judgment can be improved

Topics

1. Relation between common drivers and linear correlation
2. Measuring and comparing tail dependency

1. Relation Between Common Drivers and Linear Correlation

Quiz

- We have two portfolios.
- The distribution of annual un-inflated losses from each policy in each portfolio is assumed to have a lognormal distribution with
  - Mean 1,000
  - 99.9th percentile 50,000
  - Standard deviation 4,742
- All the policies in both portfolios are affected by only one common driver which acts on them in a similar way to that of inflation
  - The value of this "inflation" factor is unknown
  - The force of "inflation" is assumed to be normally distributed with mean 5% and standard deviation 6%
  - The force of "inflation" is assumed to be independent of the size of the loss
- Have a guess about the correlation between the losses in these two portfolios

Quiz

- Assume we had 10 years of historical loss data for the two portfolios
  - Estimation error is very large
- Expert judgment
  - High, medium, low correlation
  - How do we decide whether it is low, medium or high?
  - What value?
  - What dependency structure?

Definitions and Reminder

Total variance, un-diversifiable, diversifiable

- Reminder $$\underbrace{V[X]}_{\text{Total variance}} = \underbrace{V[E[X \mid \Theta]]}_{\text{un-diversifiable}} + \underbrace{E[V[X \mid \Theta]]}_{\text{diversifiable}}$$
- The first term can be interpreted as the un-diversifiable part of the variance and the second as the diversifiable
- Let $X_i \mid \Theta$ be identical and independent random variables which depend on the random variable Θ
- The variable Θ could be any common driver of the random variables, e.g. inflation, catastrophic event, etc.

Some Results

- The covariance of these two variables is

$$Cov(X_i, X_j) = Cov(E[X_i \mid \Theta], E[X_j \mid \Theta]) + E[Cov(X_i \mid \Theta, X_j \mid \Theta)]$$

- But $X_i \mid \Theta, \quad X_j \mid \Theta$ are i.i.d. random variables

$$= Cov(E[X_i \mid \Theta], E[X_j \mid \Theta]) = V[E[X_i \mid \Theta]]$$

- The correlation between these variables is

$$Cor(X_i, X_j) = \frac{Cov(X_i, X_j)}{\sqrt{V[X_i]V[X_j]}} = \frac{V[E[X_i \mid \Theta]]}{V[X_i]} = \frac{V[E[X_i \mid \Theta]]}{V[E[X_i \mid \Theta]] + E[V[X_i \mid \Theta]]}$$

- In this case the correlation can be interpreted as the proportion of the un-diversifiable variance
- This formula can help us estimate the correlation

Quiz: “back of the envelope calculation”

- For one policy
- Diversifiable standard deviation is *very roughly* 4,742
- Un-diversifiable standard deviation is *very roughly* 6%x1,000=60
- Diversifiable variance is *very roughly* 4,742^2
- Un-diversifiable variance is *very roughly* 60^2
- Correlation between two policies (not between portfolios) is *roughly* 0.016%

$$\approx \frac{60^2}{60^2 + 4{,}742^2} = 0.00016 = 0.016\%$$

- This compares with the exact correlation of 0.017%

Variance of Portfolio

- Assume that the $X_i s$ represent the losses from a policy i and Θ the common driver.
- Consider the losses $\sum_{i=1}^{n} X_i$ of a portfolio of n policies. The variance of the portfolio is

$$V[\sum_{i=1}^{n} X_i] = V[E[\sum_{i=1}^{n} X_i \mid \Theta]] + E[V[\sum_{i=1}^{n} X_i \mid \Theta]] = n^2 V[E[X_i \mid \Theta]] + nE[V[X_i \mid \Theta]]$$

- The standard deviation per policy is

$$\frac{\sqrt{V\left[\sum_{i=1}^{n} X_i\right]}}{n} = \frac{\sqrt{n^2 V[E[X_i \mid \Theta]] + nE[V[X_i \mid \Theta]]}}{n} = \sqrt{V[E[X_i \mid \Theta]] + \frac{E[V[X_i \mid \Theta]]}{n}}$$

- Note that the st.dev. per policy decreases as the portfolio increases and tends to the square root of the covariance as the number of policies goes to infinity

Portfolios of n Policies

- Assume that the $X_i s$ represent the losses from the i-th policy and Θ the common driver.
- Consider the losses of two portfolios $\sum_{i=1}^{n} X_i$ of n policies each.
- Applying similar calculations as in previous slides you can show that

$$Cor\left(\sum_{i=1}^{n} X_i, \sum_{j=n+1}^{2n} X_j\right) = \frac{V[E[X_i \mid \Theta]]}{V[E[X_i \mid \Theta]] + \dfrac{E[V[X_i \mid \Theta]]}{n}}$$

- Note that in this case the correlation increases as the size of portfolio n increases and $\lim_{n \to \infty} Cor\left(\sum_{i=1}^{n} X_i, \sum_{j=n+1}^{2n} X_j\right) = 1$
- This is because the diversifiable process risk is diversified away and the un-diversifiable part of the variance increases as proportion of the total variance

Quiz: “back of the envelope calculation”

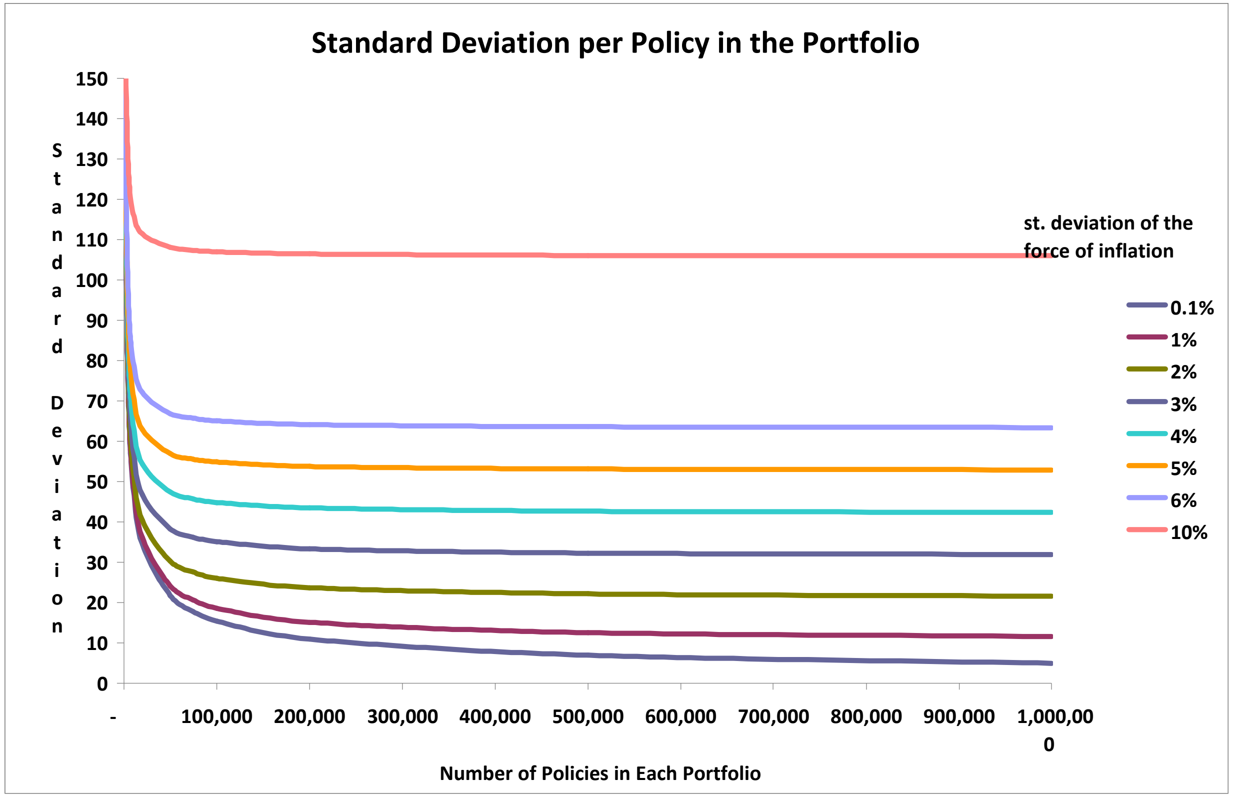
- For a portfolio of 5,000
- Diversifiable variance per policy is *very roughly* 4,742^2/5,000
- Un-diversifiable variance per policy is *very roughly* 60^2
- Correlation between two policies (not between portfolios) is *roughly* 44.5%

$$\approx \frac{60^2}{60^2 + \dfrac{4{,}742^2}{5{,}000}} = 0.445 = 44.5\%$$

- This compares with the exact correlation of 45.8%

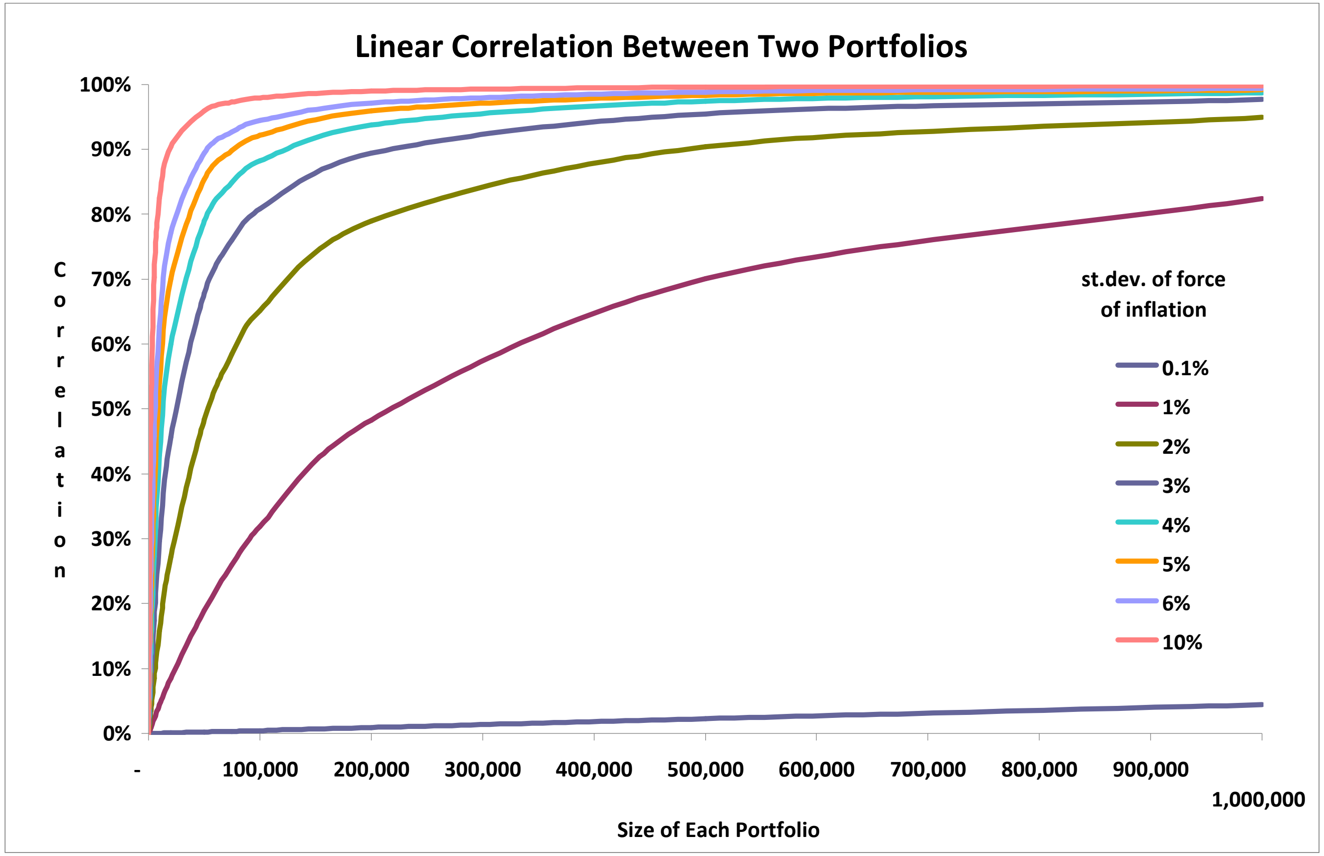
Quiz: Numerical Results
Standard Deviation per Policy

- The following graph shows the st.dev. per policy for different sizes of portfolio and values of the st.dev. of the force of inflation
- St.dev. per policy decreases but not according to the square root law and tends to a fixed value of the un-diversifiable risk



Quiz: Numerical Results
Linear Correlations

- The correlations between two portfolios of equal size are shown in the graph



Portfolios of Different Sizes

- In the more general case where one portfolio has n policies and the other m policies, the correlation is

$$Cor\left(\sum_{i=1}^{n} X_i, \sum_{j=n+1}^{n+m} X_j\right) = \frac{V[E[X_i \mid \Theta]]}{\sqrt{\left(V[E[X_i \mid \Theta]] + \frac{E[V[X_i \mid \Theta]]}{n}\right)\left(V[E[X_i \mid \Theta]] + \frac{E[V[X_i \mid \Theta]]}{m}\right)}}$$

- which again increases as the sizes of the portfolios increase and tends to 1 as n and m tend to infinity

Example 1 (as in quiz)

- Consider a portfolio of policies
- The losses of each policy follow a lognormal distribution.
  - with mean of 1,000 and
  - 99.9th percentile 50,000
- The policies are affected by a common factor which for convenience we call “inflation”.
  - The force of “inflation” d is assumed to follow a normal distribution.
  - The “inflation” factor has a lognormal distribution.
  - The losses of a policy is denoted by $X \cdot e^{d}$
    - where X follows a lognormal distribution.
- We assume that
  - the policies are not affected by any other common factor.
  - given the value of the common factor the losses of each policy are independent and identically distributed
  - the force of inflation and the size of losses are independent

Example 1

- The formulae for the variance and correlation of portfolios in the earlier slides apply and for this example it can be shown that

$$V[E[X_i | d]] = E[X]^2 \cdot V[e^d]$$

$$E[V[X_i | d]] = V[X] \cdot E[e^d]$$

where d is the force of inflation

Example 1: Correlations between portfolios of different sizes

- The force of inflation is assumed to have mean 5% and st.dev. 6% and follow a Normal distribution
- The following table shows the correlation between losses of portfolios of different sizes

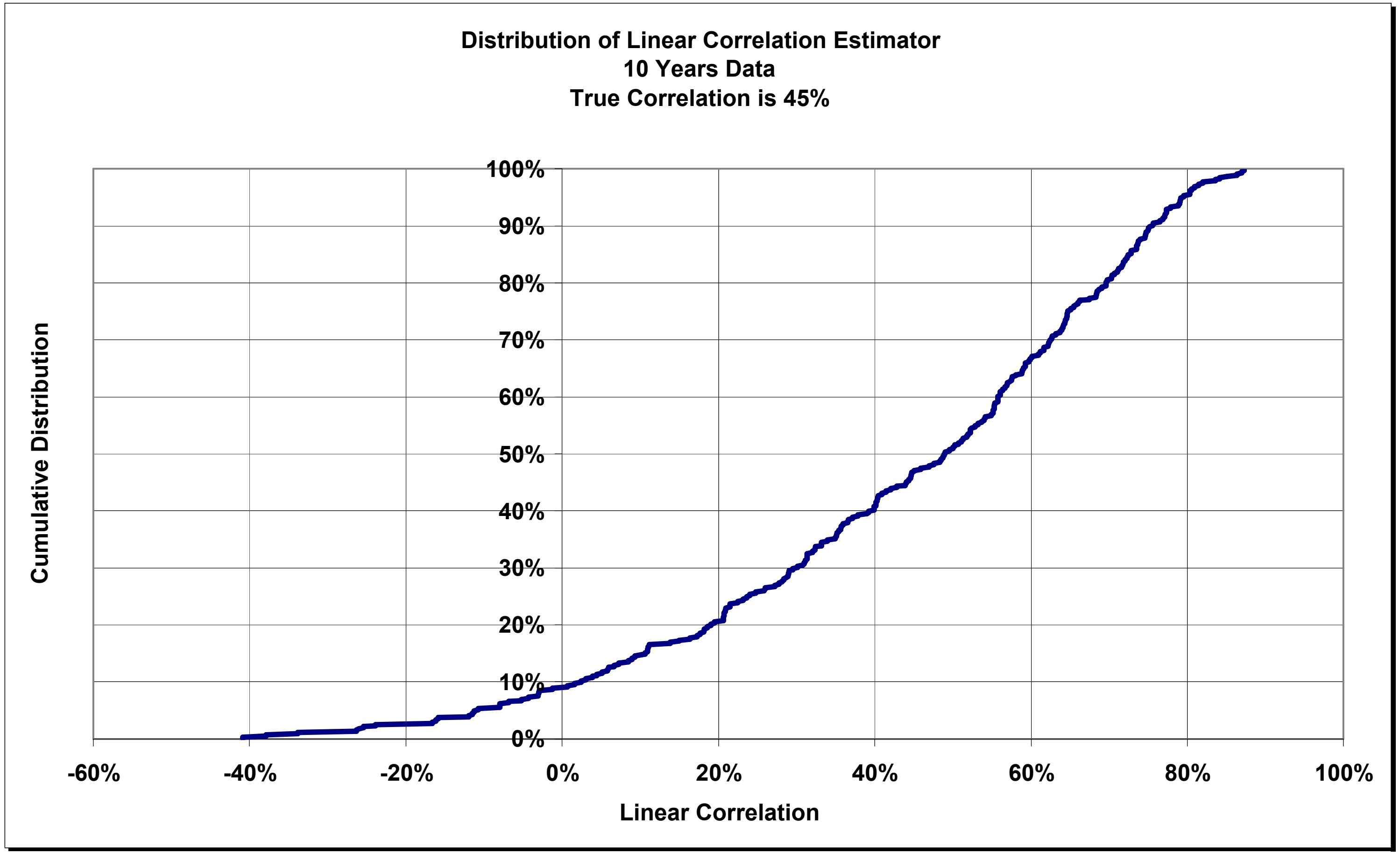
	1	100	1,000	5,000	10,000	100,000	1,000,000
1	0.0%	0.2%	0.5%	0.9%	1.0%	1.3%	1.3%
100	0.2%	1.7%	4.9%	8.7%	10.2%	12.5%	12.9%
1,000	0.5%	4.9%	14.5%	25.7%	30.1%	36.9%	37.9%
5,000	0.9%	8.7%	25.7%	45.8%	53.6%	65.7%	67.5%
10,000	1.0%	10.2%	30.1%	53.6%	62.8%	77.0%	79.0%
100,000	1.3%	12.5%	36.9%	65.7%	77.0%	94.4%	96.9%
1,000,000	1.3%	12.9%	37.9%	67.5%	79.0%	96.9%	99.4%

Linear Correlation Estimator Distribution

- Historical data would not have necessarily been of great help
- In our example, if
  - portfolios consisted of 5,000 policies each
  - the st. dev. of the force of inflation was 6%
- Then correlation would be 45%
- If we had 10 years data the distribution of the linear estimator is shown in the next slide
- The range is too wide to be of any use

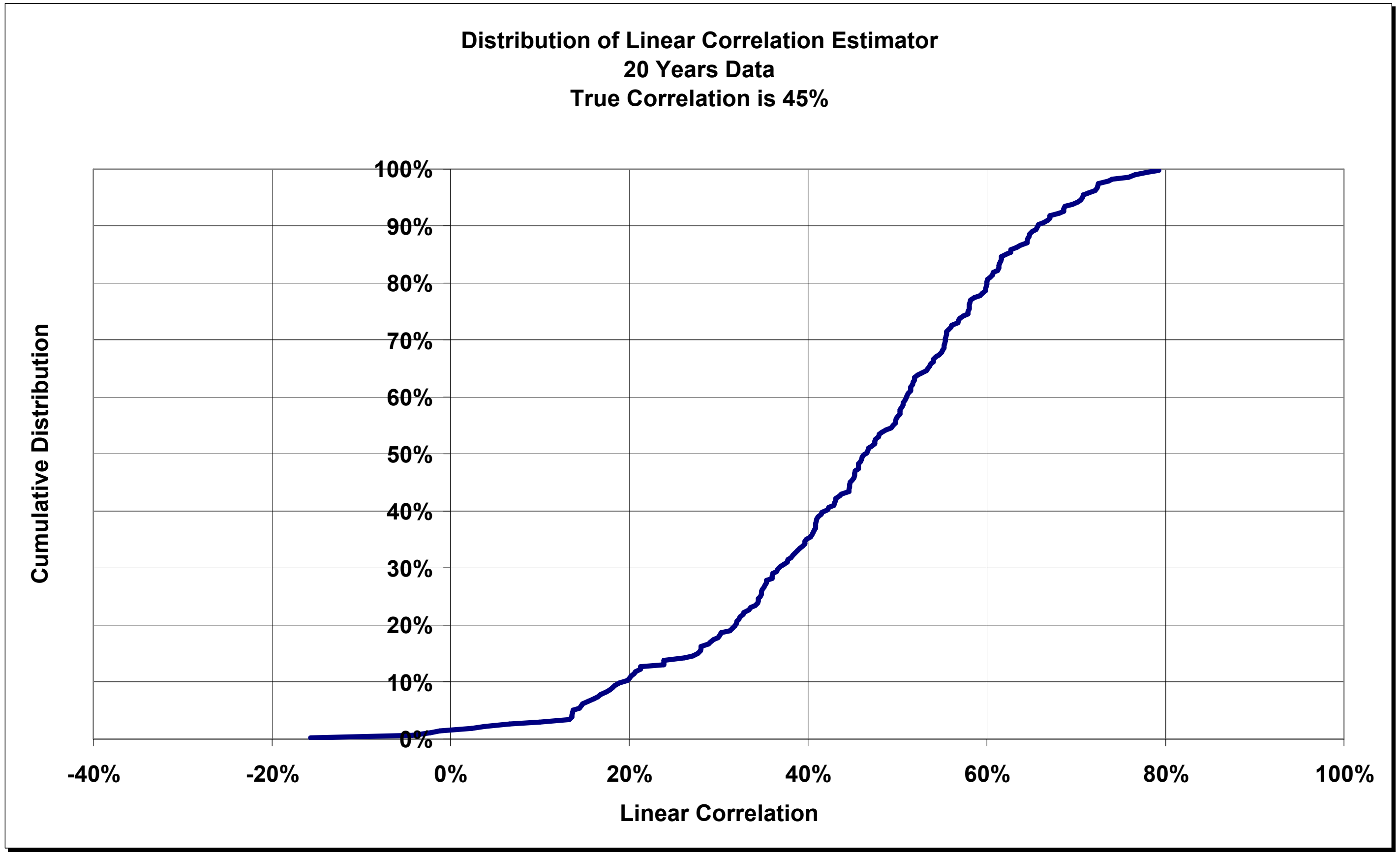
Linear Correlation Estimator Distribution

- 10 Years' data



Linear Correlation Estimator Distribution

- 20 Years' data
- The range is narrower, but it does not look a lot better



Reinsurance

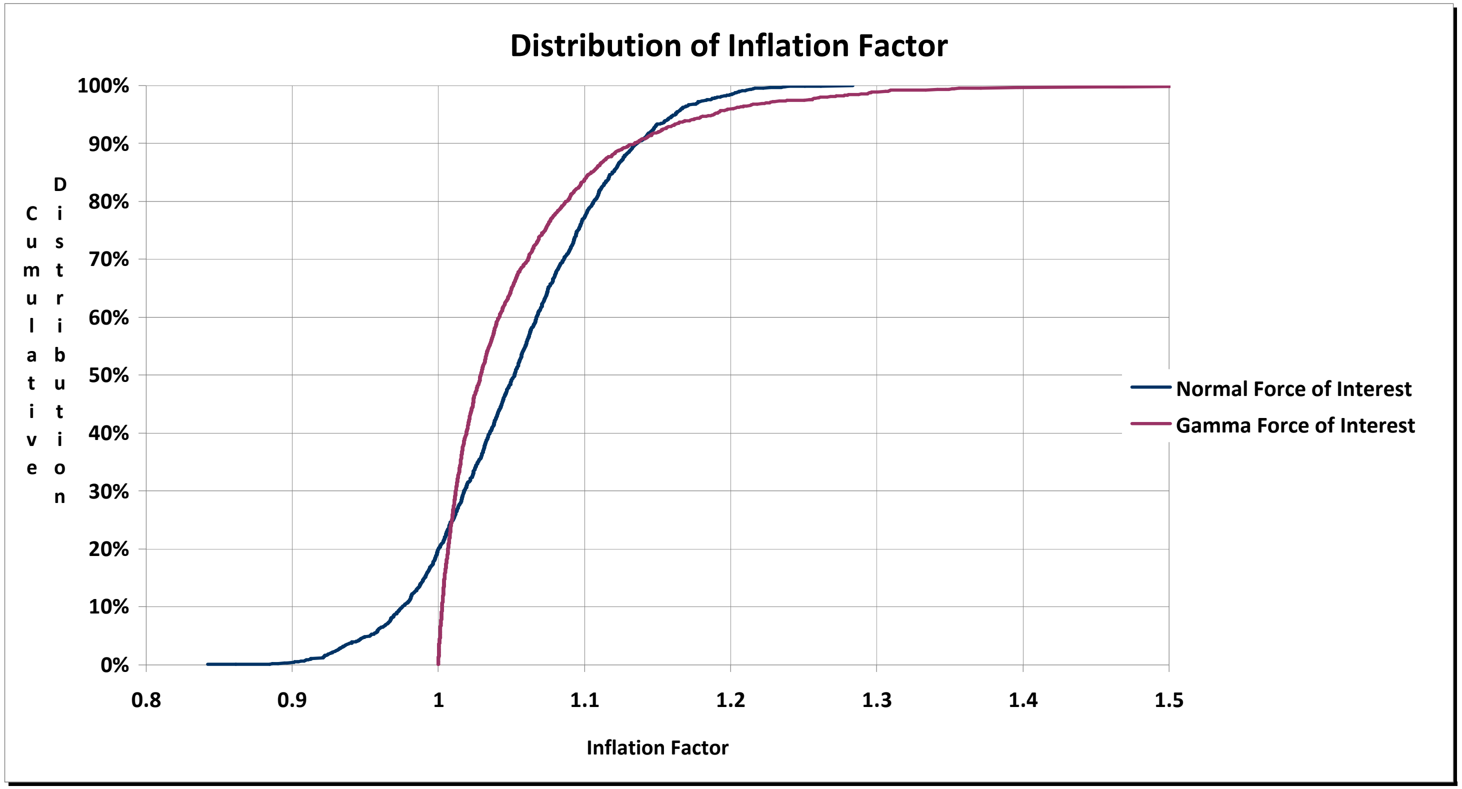
- In our example assume that
  - each of the portfolios consist of 5,000 policies
  - The st.dev. of inflation is 6% and normally distributed
  - Losses in excess of 25,000 (99.6%-ile) are covered by reinsurance with no additional premiums, reinstatements, etc.
- Ignoring reinsurance the correlation is around 45%
- With the above reinsurance in place, correlation is around 67%
- In this case reinsurance reduces the diversifiable component of the variance and correlation increases
- If “inflation” was the common driver for large losses only, then the effect of reinsurance may be reduction in the correlation
- Correlations of gross and net losses may differ
- The effect of reinsurance depends on whether increases or decreases the proportion of the diversifiable component of the variance

Example 2: Dependency Structures

- The correlations have been estimated, but nothing has been said about the dependency structure in this example.
- The dependency structure depends on the distribution of the common driver. The features of the distribution of the common factor are transformed into dependencies.
- For the numerical examples we assume that
  - the force of inflation has mean of 5% and standard deviation of 6%
  - The size of the two portfolios are assumed to be of 50,000 policies each.
- We consider two distributions for the force of inflation
  - A Normal distribution and
  - A Gamma distribution
  - They both have the same mean and standard deviation

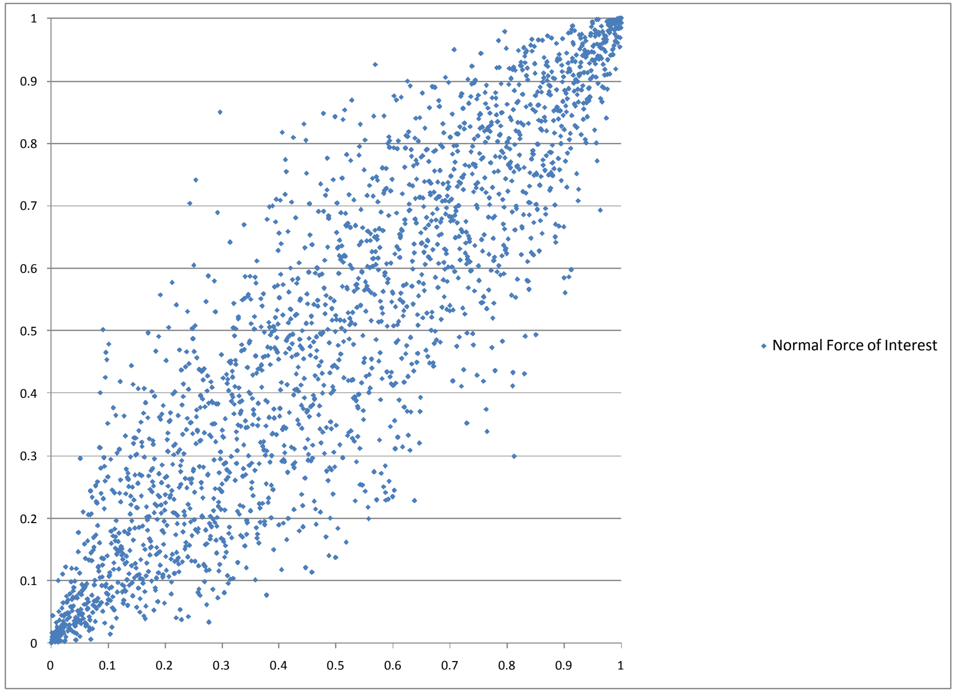
Example 2: The distribution of the inflation factor (exp(d))

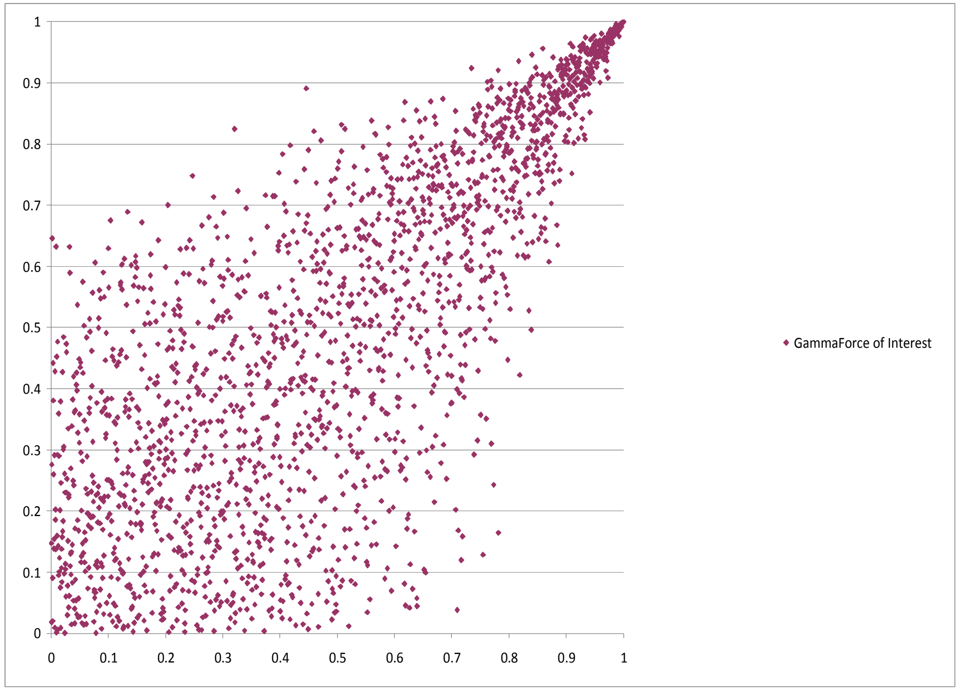
- Both distributions have force of inflation with mean 5% and st.dev. 6%



Example 2: The dependency structures differ significantly for the two distributional assumptions

- The dependency structure is symmetric for normal force of inflation and has stronger upper tail dependency for lognormal force of inflation
- Both copulas have very similar linear correlations of about 90%





More than One Common Drivers

- In reality, different portfolios or different lines of business are not affected only by a single common factor
- Policies within the same line of business may have their own common drivers which affect only this line of business
- In addition to those drivers there may be common drivers which affect more than one line of business
- The interpretation of the linear correlation as the percentage of the un-diversifiable variance out of the total variance still applies
  - Demonstrated by the following example
  - Similar results hold for other types of common driver

Example 3: Two portfolios each with its own common drivers and a common driver affecting both portfolios

- Portfolio A: Losses X_i from each policy
  - have mean μ_{A1} and variance σ^2_{A1}
  - a common driver affects all policies of portfolio A only. The effect of this driver is that the mean of each policy is the same for all policies and becomes a r.v. with mean μ_{A2} and variance σ^2_{A2}
- Portfolio B: Losses Y_i from each policy
  - have mean μ_{B1} and variance σ^2_{B1}
  - a common driver affects all policies of portfolio B only. The effect of this driver is that the mean of each policy is the same for all policies and becomes a r.v. with mean μ_{B2} and variance σ^2_{B2}
- A driver Θ affects policies of both portfolios and the losses for portfolio A and B become $\Theta \cdot X_i$ and $\Theta \cdot Y_i$ respectively

Example 3: Two portfolios each with its own common drivers and a common driver affecting both portfolios

PORTFOLIO A

$$E[X_i] = \mu_{A1} \quad V[X_i] = \sigma^2_{A1}$$

Driver affecting only policies in portfolio A makes mean a random variable

$$E[\mu_{A1}] = \mu_{A2}$$

$$V[\mu_{A1}] = \sigma^2_{A2}$$

PORTFOLIO B

$$E[Y_j] = \mu_{B1} \quad V[Y_j] = \sigma^2_{B1}$$

Driver affecting only policies in portfolio B makes mean a random variable

$$E[\mu_{B1}] = \mu_{B2}$$

$$V[\mu_{B1}] = \sigma^2_{B2}$$

ΘX_i

Driver Θ affecting both portfolio A and portfolio B

ΘY_j

Example 3 Results: Two portfolios each with its own common drivers and a common driver affecting both portfolios

- If Portfolio A has n policies and portfolio B has m policies, after some calculations similar to those in earlier slides it can be shown that the correlation between the two portfolios is
- Note that the correlation does not tend to 1 as n and m go to infinity

$$Cor\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \frac{V[\Theta]}{\sqrt{\left(V[\Theta] + E[\Theta^2]\frac{\sigma_{A2}^2}{\mu_{A2}^2} + E[\Theta^2]\frac{\sigma_{A1}^2}{n \cdot \mu_{A2}^2}\right)\underbrace{\left(V[\Theta] + E[\Theta^2]\frac{\sigma_{B2}^2}{\mu_{B2}^2} + E[\Theta^2]\frac{\sigma_{B1}^2}{m \cdot \mu_{B2}^2}\right)}_{\text{Similarly for Portfolio B}}}}$$

Un-diversifiable variance due to the common driver affecting both portfolios

Un-diversifiable variance due to the driver common to the policies in portfolio A only

Diversifiable process variance of portfolio A

Numerical Example 3: Two portfolios each with its own common drivers and a common driver affecting both portfolios

PORTFOLIO A
(10,000 policies)

$E[X_i] = \mu_{A1}$

$V[X_i] = \sigma_{A1}^2 = 4{,}740^2$

Driver affecting only policies in portfolio A makes mean a random variable

$E[\mu_{A1}] = \mu_{A2} = 1{,}000$

$V[\mu_{A1}] = \sigma_{A2}^2 = 100^2$

PORTFOLIO B
(10,000 policies)

$E[Y_j] = \mu_{B1}$

$V[Y_j] = \sigma_{B1}^2 = 4{,}740^2$

Driver affecting only policies in portfolio B makes mean a random variable

$E[\mu_{B1}] = \mu_{B2} = 1{,}000$

$V[\mu_{B1}] = \sigma_{B2}^2 = 150^2$

Driver Θ affecting both portfolio A and portfolio B

$E[\Theta] = 1$

$V[\Theta] = 0.06^2$

ΘX_i

ΘY_j

Example 3 Results: Two portfolios each with its own common drivers and a common driver affecting both portfolios

- Substituting the numbers in the following formula

$$Cor\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \frac{V[\Theta]}{\sqrt{\left(V[\Theta]+E[\Theta^2]\frac{\sigma_{A2}^2}{\mu_{A2}^2}+E[\Theta^2]\frac{\sigma_{A1}^2}{n\cdot\mu_{A2}^2}\right)\left(V[\Theta]+E[\Theta^2]\frac{\sigma_{B2}^2}{\mu_{B2}^2}+E[\Theta^2]\frac{\sigma_{B1}^2}{m\cdot\mu_{B2}^2}\right)}}$$

$$\frac{0.06^2}{\sqrt{\left(0.06^2+1.0036\frac{100^2}{1{,}000^2}+1.0036\frac{4{,}740^2}{10^4\cdot 1{,}000^2}\right)\left(0.06^2+1.0036\frac{150^2}{1{,}000^2}+1.0036\frac{4{,}740^2}{10^4\cdot 1{,}000^2}\right)}} = 17\%$$

- Have a guess if n=m=infinity
- Correlation = 19%
- Note that we have not said anything about the distributions. The linear correlation here depends only on the first two moments of the distributions involved and the size of the portfolio

Relation Between Common Drivers and Linear Correlation

Conclusion

- Linear correlations are about the contributions of the different sources of volatility to the total volatility (variance)
- The subjective expert judgment may be assisted if we
  - Identify all the common drivers
  - Estimate the contribution of each of them to the variance of the portfolios
  - Estimate the diversifiable part of the volatility by taking to account the size of the portfolios
  - Compare the variance due to the driver common to both portfolios to the other components of the variance
- Note that the above discussion applies to linear correlation
- The distribution of the common driver(s) affect the dependency structure and tail dependency.

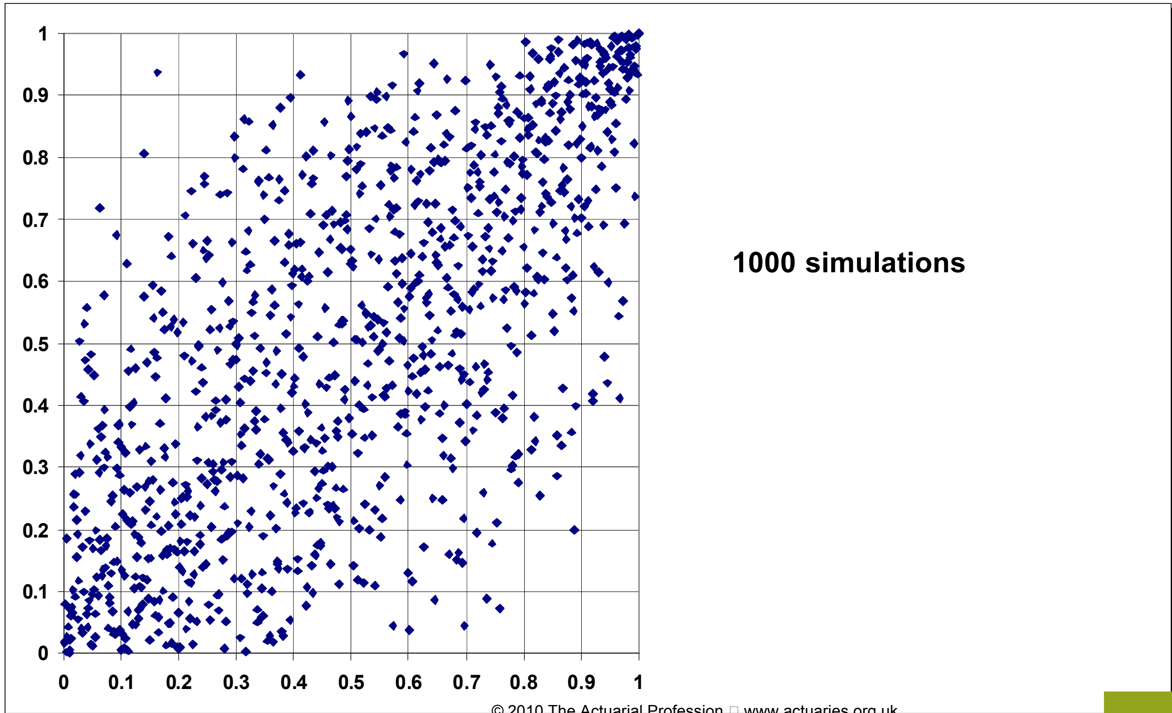
2. Measuring and comparing tail dependencies

Alternative Ways of Looking at Dependencies

- How the mass of probability at the tail of a dependency structure compares with the mass of probability at the tail when assuming independence?
- Tail dependence: Given that a loss is higher than the 1 in 100 for one LoB what is the probability of a loss higher than the 1 in 100 in another LoB?
- Given a dependency structure and given that we have the 1 in 100 loss for one LoB, how much higher is the probability of having a loss greater than the 1 in 100 in another Lob compared to that assuming independence?

Increase in Mass of Probability in a Quadrant

- If random variables were independent 1% of dots (10 dots) would be in the top right square (subject to simulation error)
- We can count more than 10 dots in the top right square



Increase in Mass of Probability in a Quadrant – Theoretical Answer

- What is the probability that r.v. X will be higher than the u(*100)-th percentile and variable Y higher than the v(*100)-th percentile
- If independent then (1-u)*(1-v)
- If dependency is given by the copula C(u,v), then

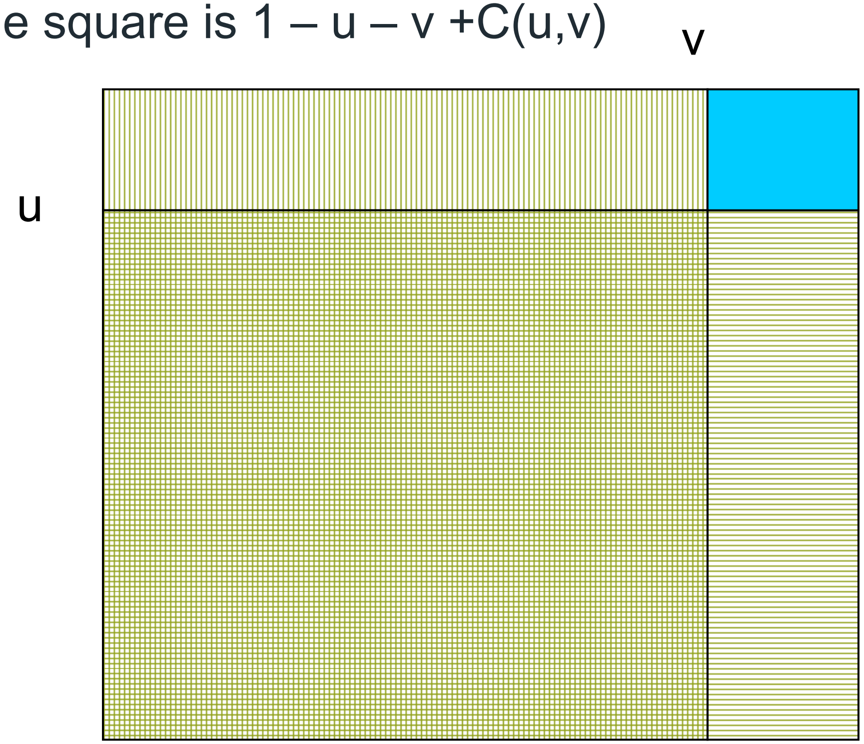
$$1-(C(u,1)+C(1,v)-C(u,v)) = 1-u-v+C(u,v)$$

- Ratio of the two is

$$(1-u-v+C(u,v))/((1-u)(1-v))$$

Increase in Mass of Probability in a Quadrant – Theoretical Answer Explanation

- The probability in the whole square is 1
- The probability to the left of v is v
- The probability below u is u
- The probability in the area excluding the blue square is u+v-C(u,v) because we have double counted the probability in the square with the crossed lines
- The probability in the blue square is 1 – u – v +C(u,v)

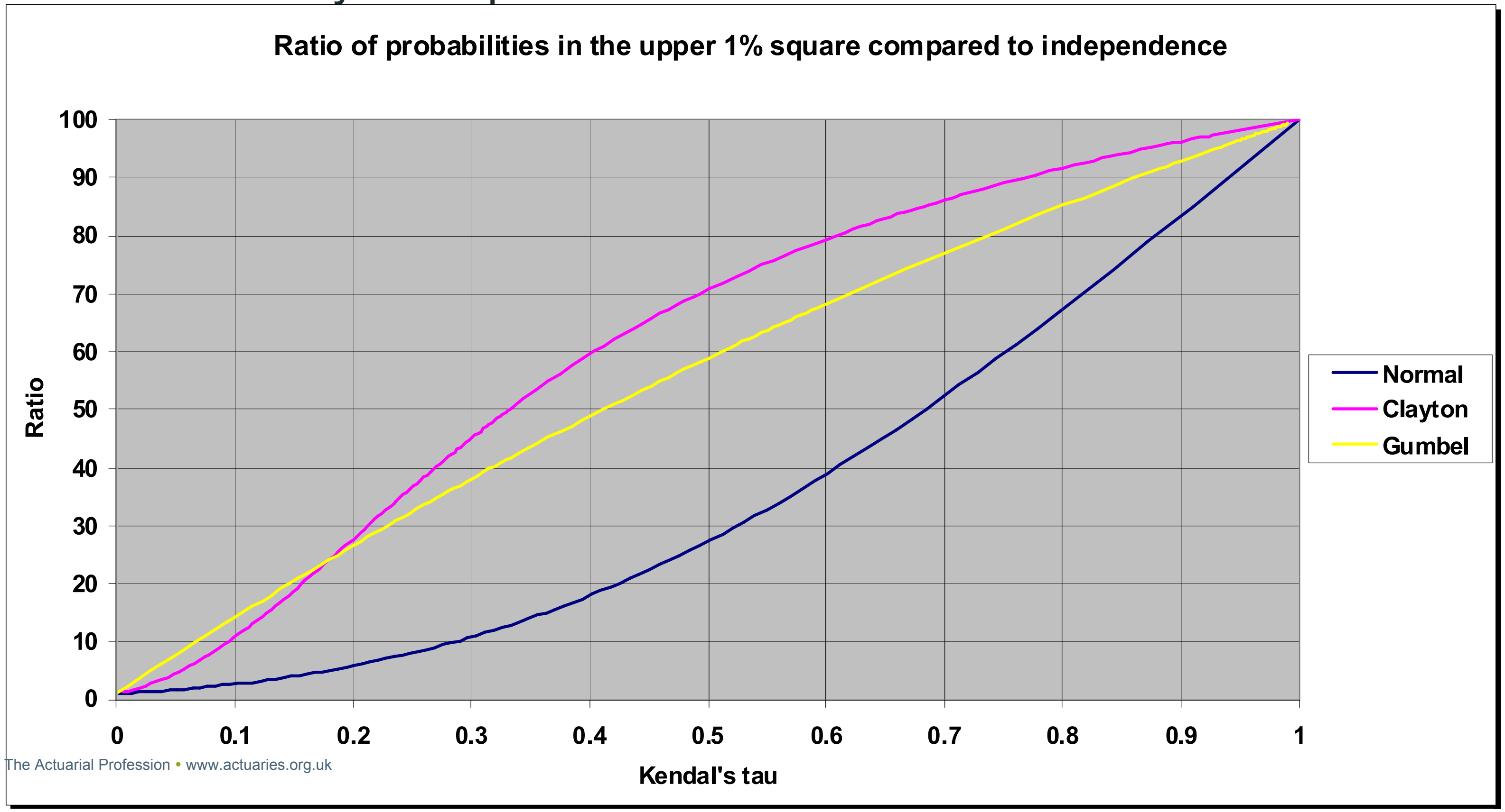


Increase in Mass of Probability in a Quadrant

- We could assume that u=v
- For example the 1 in 100 events: u=v=0.99
- If the two r.v. were independent then the probability that both exceed their 99-th percentile is 0.01x0.01=0.0001
- If the two r.v. were fully dependent then the probability that both exceed their 99-th percentile is 0.01
- As the dependence increases the probability in the top right 1% quadrant will increase from 0.0001 to 0.01. As we move from independence to full dependence the mass of probability in the top 1% quadrant will become 100 times higher.
- How fast this increase happens depends on the dependency structure

Increase in Mass of Probability in top right 1% Quadrant: Examples

- Increase is faster for Clayton and Gumbel and slower for Normal
- For example, for Kendal's tau of 0.5 the mass of probability in the top 1% quadrant increases (compared to that under independence) 28 times under Gaussian dependence, 59 times for the Gumbel copula and 71 times for the Clayton copula



Tail Factor

- The quantity we examined in the previous slides should not be confused with the coefficient of tail dependence
- We examined how many times higher the mass of "probability" at the top right quadrant of a copula is compared to the independent copula. This is given by

(1-2u+C(u,u))/((1-u)(1-u))

- The coefficient of tail dependence is defined as

$$\lim_{u \to 1^-} \Pr[Y > F_Y^{-1}(u) \mid X > F_X^{-1}(u)] = \lim_{u \to 1^-} \frac{1 - 2u + C(u,u)}{1-u}$$

- For the Normal copula this coefficient is 0 for any correlation less than 1, while it is positive for the Gumbel, t and Clayton copulas
- We did not take the limits

Measuring and comparing tail dependency
Conclusion

- The probabilities of joint events in the different quadrants of a copula could be used to
  - improve our understanding of the different copulas
  - compare different dependency structures and ensure consistency
  - help expert judgement

Things to take with you

1. Linear correlation depends on the percentage contribution of the common drivers to the total variance
   - Correlation depends on the size of the portfolios
   - The features of the distribution of the common factors translates into dependency structure
2. The probabilities of joint events in the different quadrants of a copula could be used to understand and compare copulas
3. Your personal belongings

Acknowledgements

- I would like to thank my colleagues at the FSA and especially Andy Macnair for some useful feedback.

Questions or comments?

Expressions of individual views by members of The Actuarial Profession and its staff are encouraged.

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