# Dependence elicitation for risk and decision analysis

Tim Bedford, Christoph Werner, John Quigley

Department of Management Science
Strathclyde Business School
University of Strathclyde
Glasgow, Scotland

# Abstract

When assessing uncertainty, whether for risk assessment or as part of a decision analysis, it is important to consider whether the uncertainty we wish to model for variables of interest should include stochastic dependency. Simple examples suffice to show the real significance of capturing dependency in the outputs of models. Such dependency often arises because there are factors outside the scope of the model which link uncertainties between the variables within the model.

Copulas and vines provide mathematical structures, generalizing parametric multivariate distributions, with which dependency can – in principle – be modelled. However they do not address the issue of actually quantifying a dependence structure in a specific context. In practice this has often been done by asking experts to specify bivariate correlations – which makes the implicit assumption that correlation is a quantity that experts should be good at specifying. A better, but less used approach is to ask about conditional exceedance probabilities for one variable, given an exceedance event for another: For example, the probability that X exceeds its median given that Y exceeds its median.

In this talk we shall discuss new approaches to dependence elicitation which generalise the exceedance approach. We show that multiple elicitations of exceedance probabilities can be made with exact lower and upper feasible bounds generated from previous elicitations by an LP problem. The elicitation process makes use of an approach to rationale development by the individual experts that both allows them to share understandings of the qualitative factors leading to dependence, and also allows them to provide insights to the owners and stakeholders of the broader risk management framework. This feedback is considered a critical element of risk management and is incorporated explicitly into risk management standards. This work has benefited greatly from support of the COST network IS1304.

# Topics

- Quick review of dependency models in PRA and DM
  - CCF, Copulas, Markov trees, joint normal, NORTA, vines,
- Dependency elicitation overview
- Sequential refined partitioning elicitation
- Rationales
- Conclusions

# Risk and decision analysis context

- Complex models in risk analysis
  - Fault/event tree/Markov/DFM model for system
  - Consequence models eg dispersion in air after accident
- Need to consider sensitivity and uncertainty of model outputs
- Problem of partial specification
- Typical approach to put distributions on model parameters and propagate through the model…but such distributions have to be meaningful, and this entails making them dependent models
- Another source of dependency is model incompleteness – events in the model may be dependent because we have not captured all the relevant events within the model – CCF
- CCF models typically introduce "bucket" of undefined correlating factors which lead to simultaneous, coupled or cascade failures

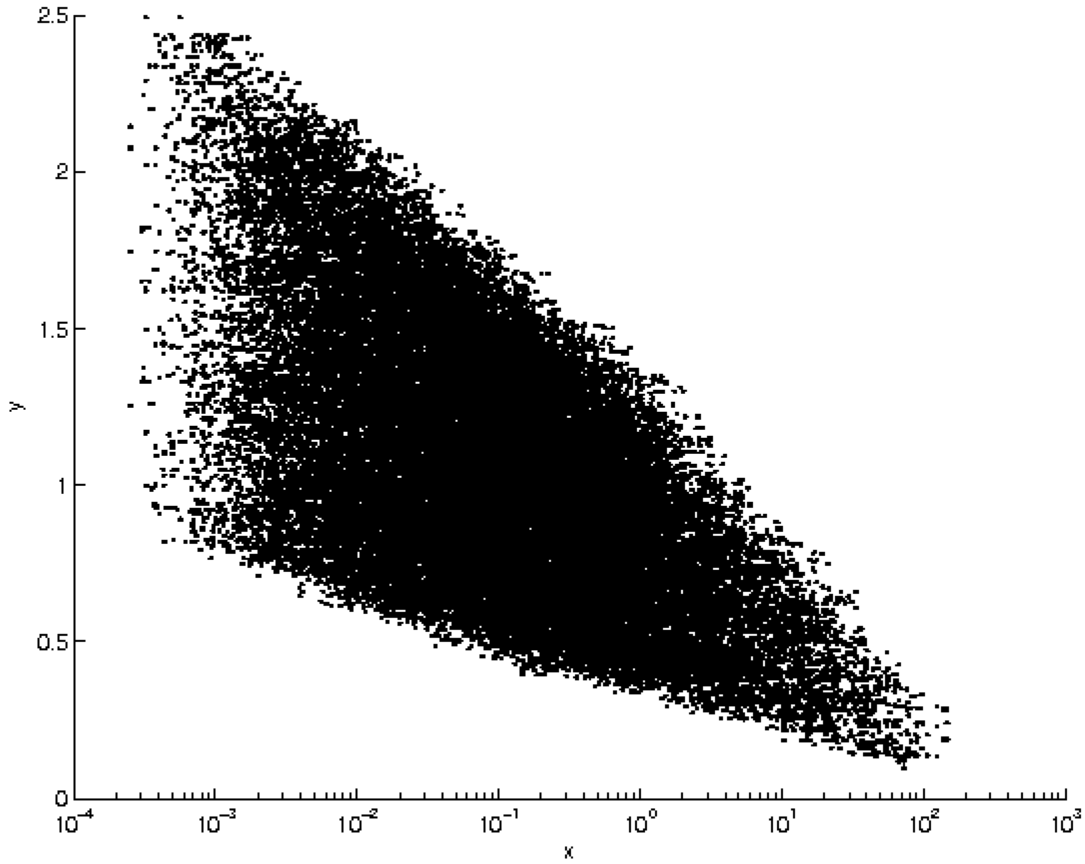
# Example – lateral plume spread

- Simple model used for diffusion of contaminant clouds from single source
- At downwind distance x, lateral spread of plume follows power law

$$\sigma_y(x) = A_y x^{B_y}$$

- Experts give judgements about several downwind plume spreads

# Scatter plot for dispersion params




Uncertainty distribution on the parameters derived from the expert judgements

Probabilistic Inversion of Expert Judgments in the Quantification of Model Uncertainty, Kraan and Bedford, *Management Science 2005*

# Expert assessment methods

- Many methods for expert assessment of distributions - for applications in reliability/risk often non-parametric
- “Traditionally” experts provide input by
  - Means, covariances..
  - Marginal quantiles, product-moment correlations
  - Marginal quantiles, rank correlations
- Consistency problems:

  Marginals, correlations

# Iman-Conover dependency method (NORTA)

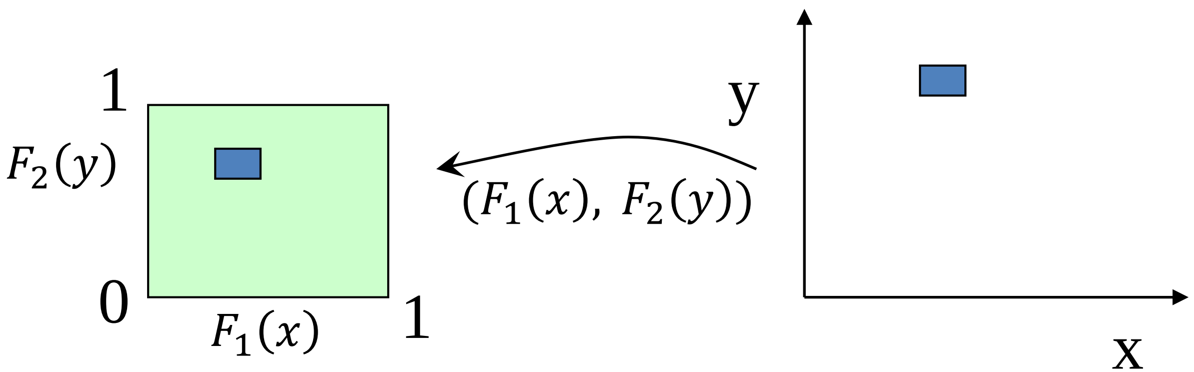
- Assumes marginal distributions known or elicited
- Idea: transform each distribution to normal
- Method (or variants) is commonly used in commercial software
- Requires input of a correlation matrix
  - Must be positive definite
  - Lots of algebraic constraints on the entries of the matrix

$$\Phi^{-1}F(X)$$

$$\begin{tikzcd}
 & (N_1, N_2) \arrow[dl] \arrow[dr] & \\
X_1 & & X_2
\end{tikzcd}$$

# Copula

- The joint distribution of the uniformised variables….




$$f_{12}(x,y) = f_1(x)f_2(y)c(F_1(x), F_2(y))$$

- Key idea is to use copula's constructively: Given a copula and marginal distributions you specify the joint distributions

# Degree of dependency …

- Many families of copula available, often with one parameter that is linked to the correlation

- Everyone has their favourite parametric copula family….
- Mine allows a lot of flexibility and use of "real world" parameters

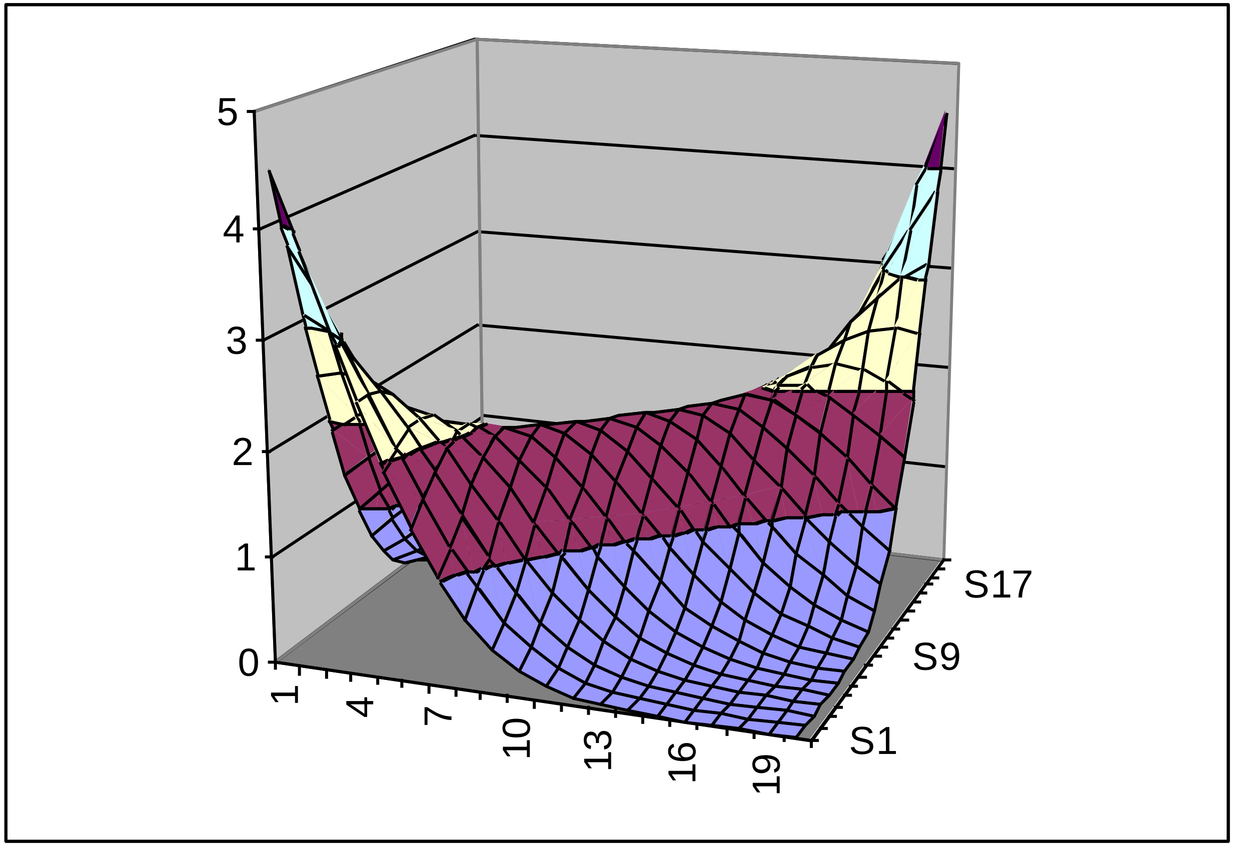
# Minimum information copulae

- Partially specify the copula, eg by (rank) correlation, or by other “observable” variables
- Find “most independent” copula given information specified
- Minimize relative information to independent copula= uniform distribution

$$I(f) = \int\int f(u,v)\log(f(u,v))dudv$$

- Min information is coordinate free criterion

# Min inf copula density with rank correlation=0.8



# What do vines do?

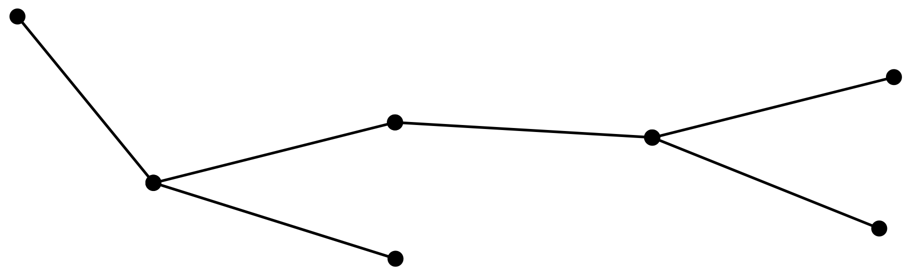
- Graphical representation of multivariate distribution
- Used when marginals are known, continuous and invertible
- Might be used in constructing a subjective distribution, or in modelling a multivariate dataset
- Extends idea of a copula to multiple dimensions
- First idea for “stacking” two dimensional copulas by Harry Joe, then Roger Cooke created graphical representation, and Bedford and Cooke gave basic theorems on existence, information etc in 2002, Ann Stat, “Vines – a new graphical model for dependent random variables”

# Markov trees

- "Patch" copulas together to build up multivariate distribution using conditional independence

- Application in particular to specification of joint distributions in uncertainty analysis.
  - (Minimum information) copulae used to couple random variables
  - Marginals specified plus certain (conditional) rank correlations
  - Main advantage is no algebraic restrictions on correlations
  - Disadvantage is difficulty of assessing correlations

# Markov tree example

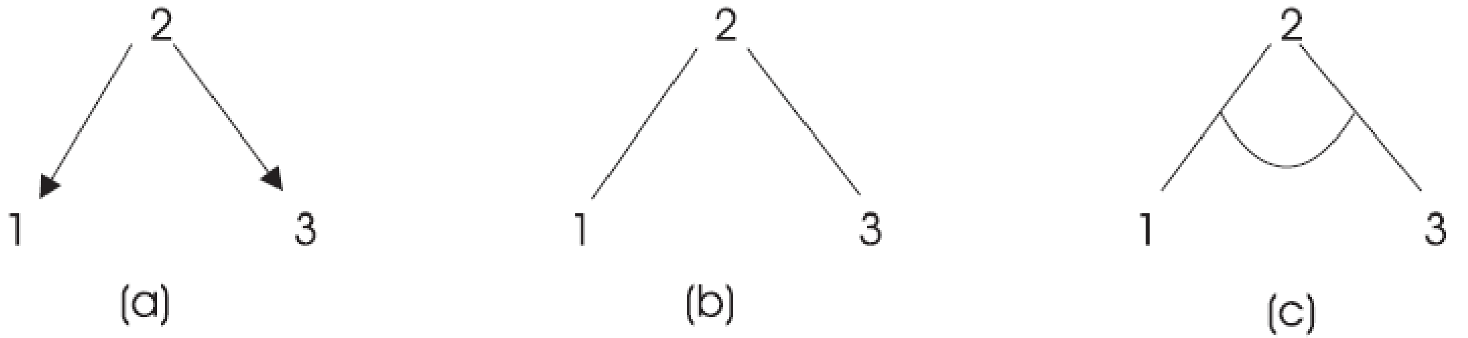


- Two variables are conditionally independent given a variable between them on the tree

Decomposition Theorem

$$f(x_1,\dots,x_n)=\frac{\prod_{(i,j)\in E} f_{ij}(x_i,x_j)}{\prod_{i\in N}(f_i(x_i))^{d(i)-1}}=f_1\dots f_n\prod c_{ij}(F_i(x_i),F_j(x_j))$$

# Extension from Markov trees to vines




Dist of 1 given 2
dist of 3 given 2,

1,3 indep given 2

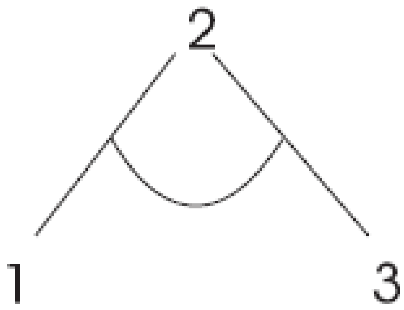
Dist of 1 and 2,
dist of 3 and 2,

1,3 indep given 2

Dist of 1 and 2,
dist of 3 and 2,

1,3 dep given 2

# A simple vine distribution...



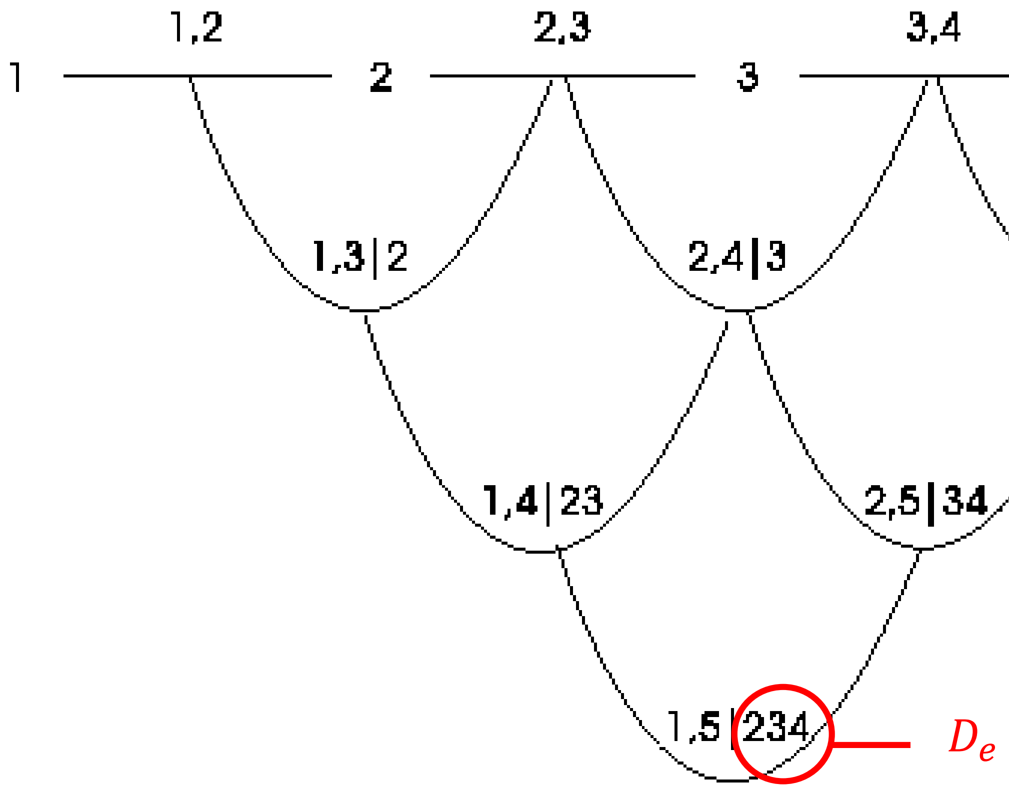

Specify marginals
Specify copulas $c_{12}, c_{23}, c_{13|2}$

Sampling procedure

- Sample $u_1$
- Sample $u_2$ using $c_{12}$ and $u_1$
- Compute conditionals $u_1 | u_2$ and $u_3 | u_2$
- Sample $u_3$ using $c_{13|2}$ and $u_1, u_2$

Joe 1997 Paired copula construction
Cooke, Bedford and Cooke, Cooke and Kurowicka

# Vine example

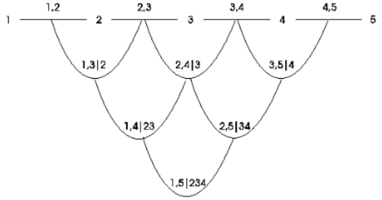



$E_4$

$T_4$

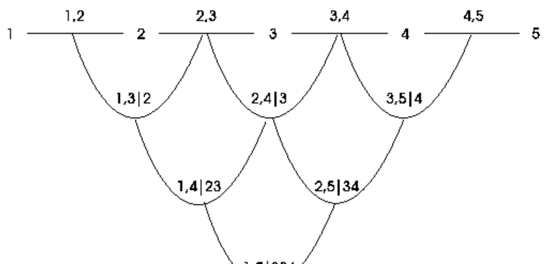
# Rank correlation vine

- Specify the rank correlation on each branch of the vine
- Any number between –1 and +1 will do
- No algebraic restrictions
- Same number of parameters as usual correlation matrix



# Partial correlation vine for normal distribution

- For multivariate normal, specify the partial (=conditional) correlation on each branch of the vine
- Any number between –1 and +1 will do
- No algebraic restrictions
- Same number of parameters as usual correlation matrix



# Min inf copula with basis functions estimated from data

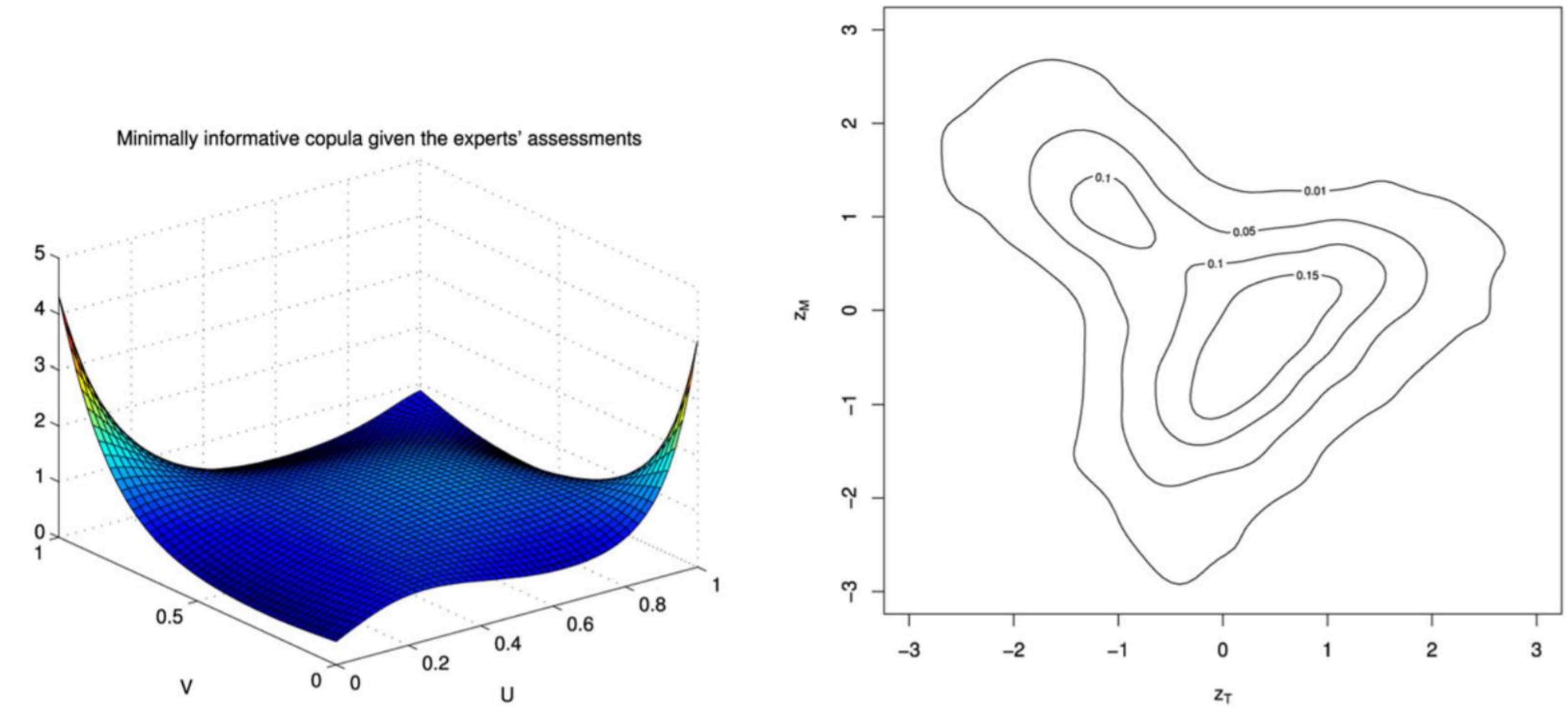
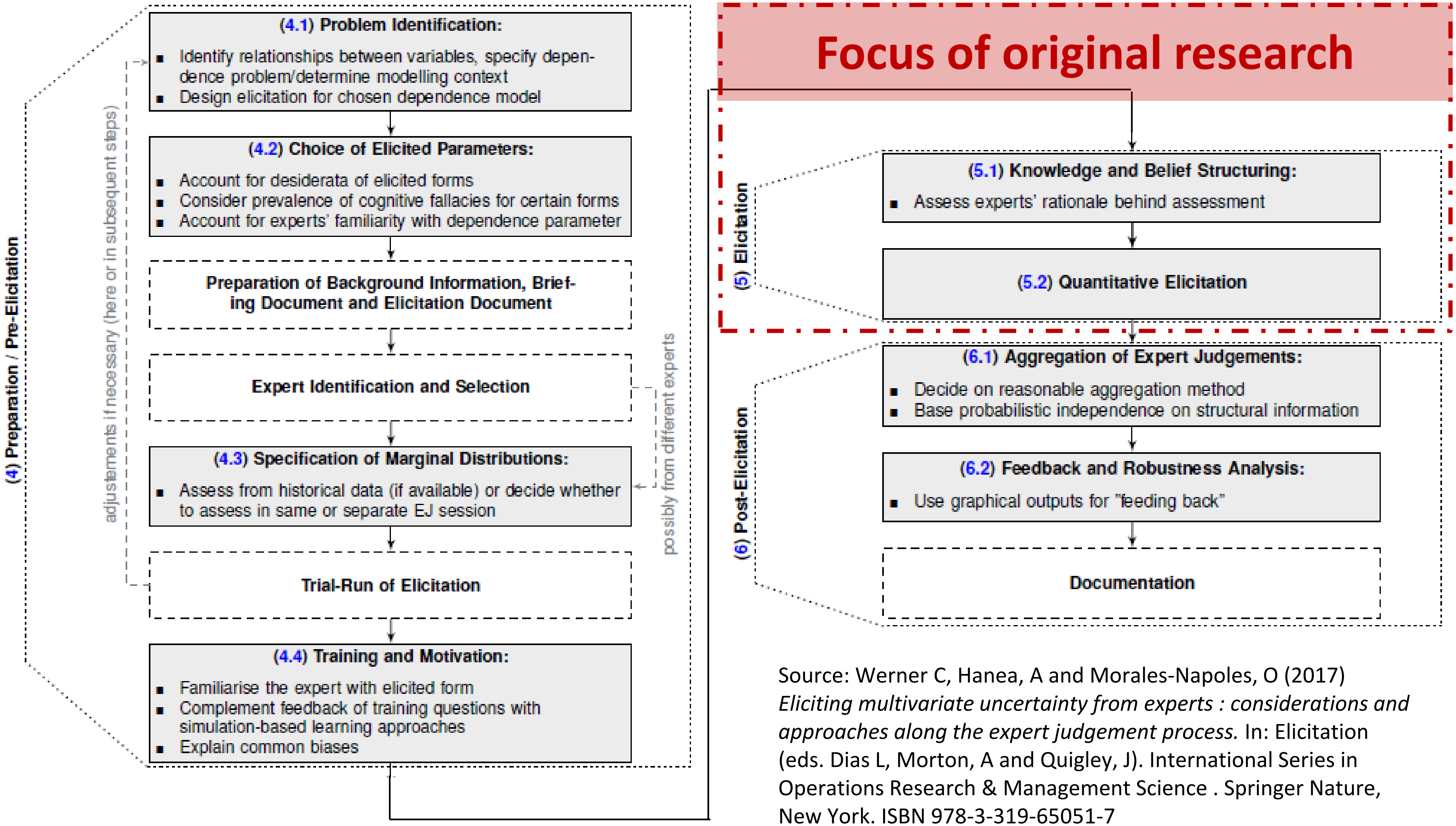



**Fig. 6.** The minimally informative copula between $T$ and $M$ and transformed contour plot, Norwegian stock data.

*Risk Analysis, Vol. 36, No. 4, 2016* DOI: 10.1111/risa.12471
Approximate Uncertainty Modeling in Risk Analysis with Vine Copulas
Tim Bedford, Alireza Daneshkhah, and Kevin J.Wilson

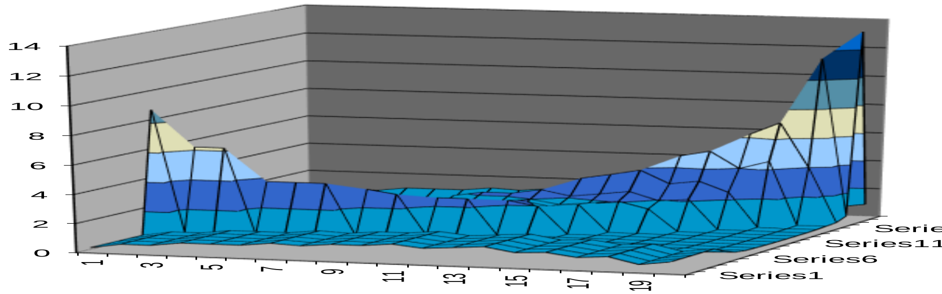
# Dependence elicitation:
## managing the overall process




Source: Werner C, Hanea, A and Morales-Napoles, O (2017) *Eliciting multivariate uncertainty from experts : considerations and approaches along the expert judgement process.* In: Elicitation (eds. Dias L, Morton, A and Quigley, J). International Series in Operations Research & Management Science . Springer Nature, New York. ISBN 978-3-319-65051-7

# Dependency elicitation using min inf

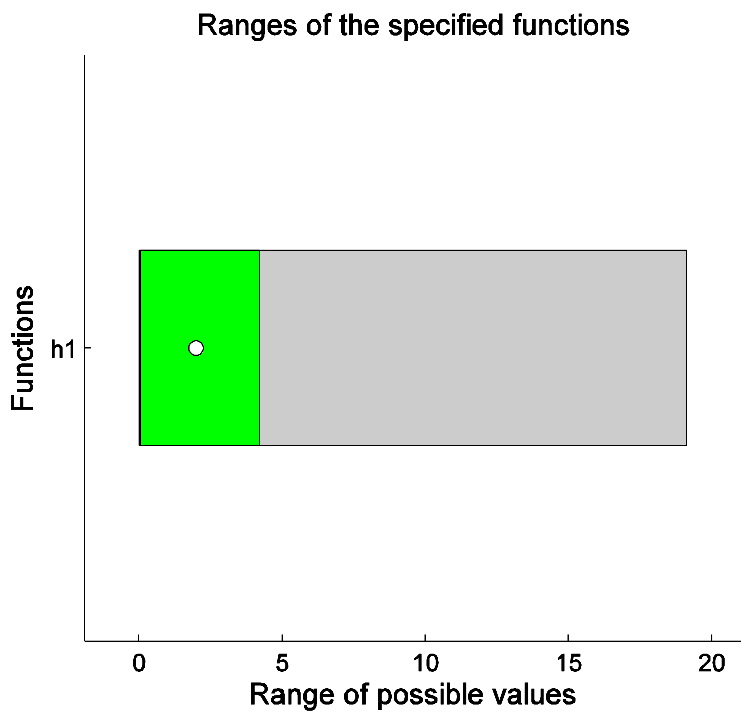
- Consider two exponential lifetimes and specify difference in observed lifetime. ie quantiles for |X-Y|
- Expert assesses
  - P(X-Y<0.3)=0.3
  - P(X-Y<0.9)=0.7

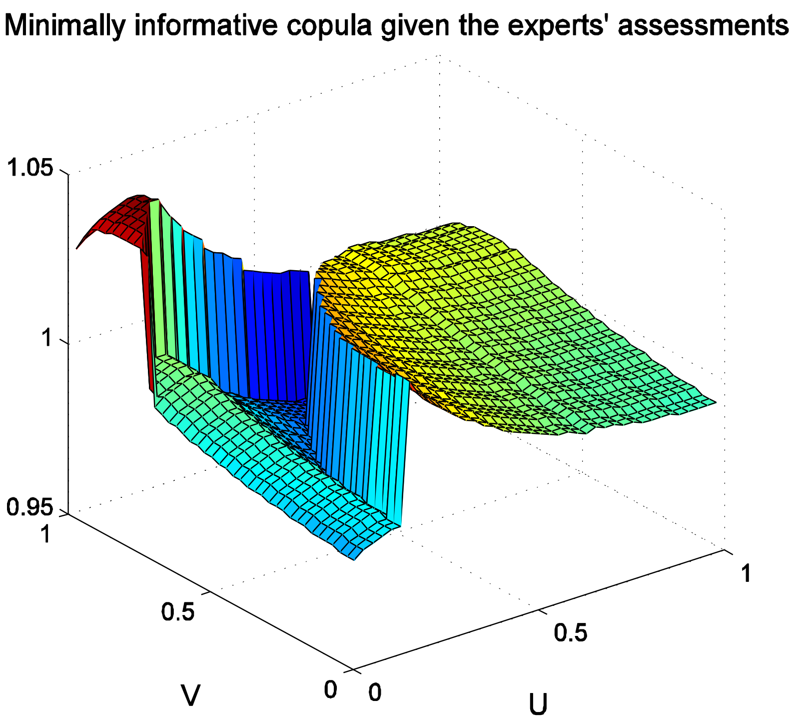



Min inf copula given the expert assessments

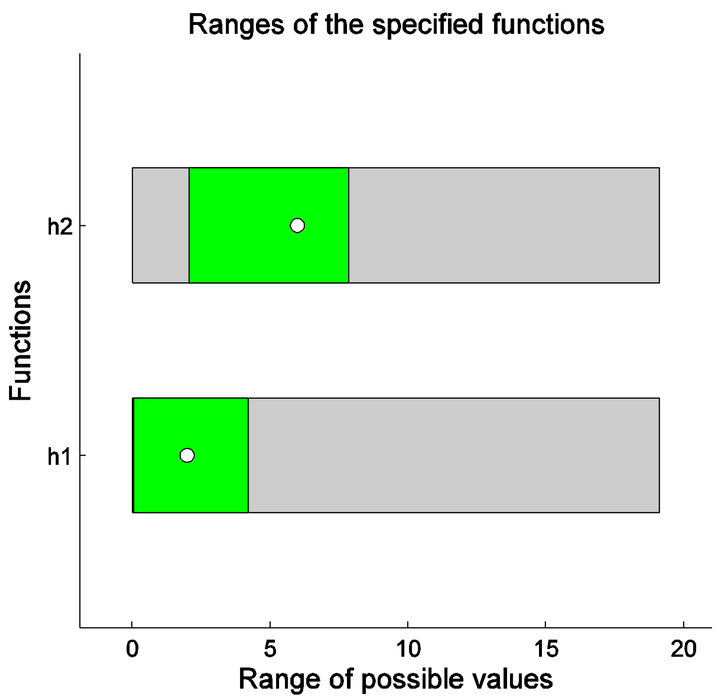
# More generally…
# Sequential Elicitation – Step 1

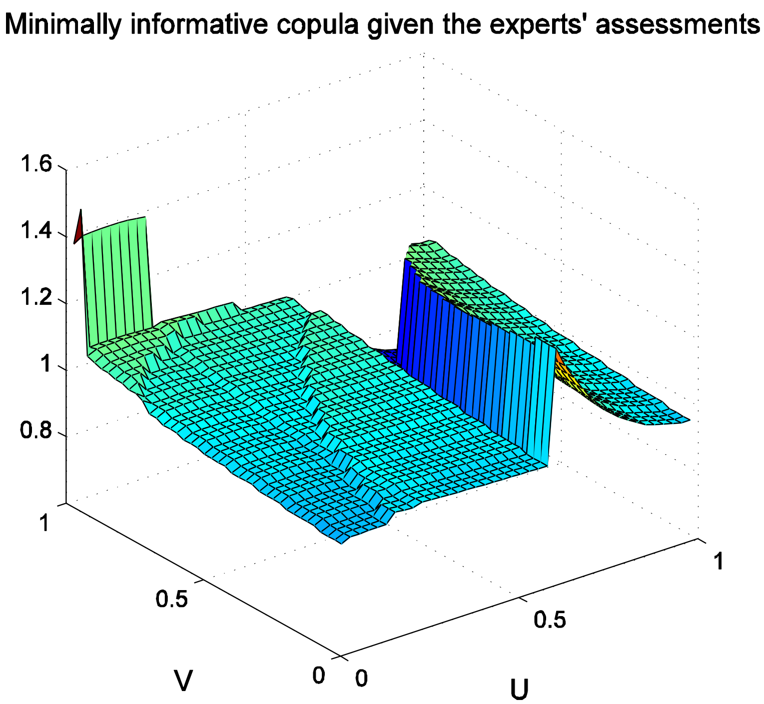
University of Strathclyde Business School



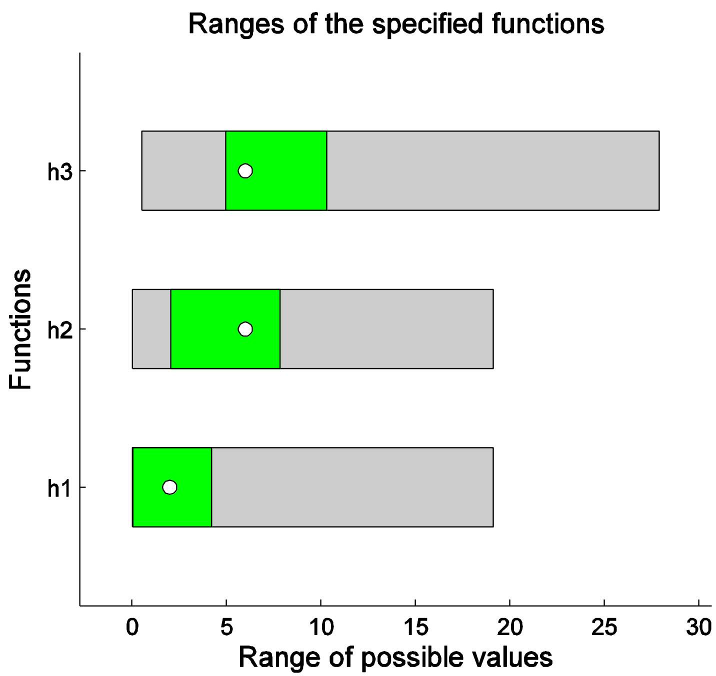



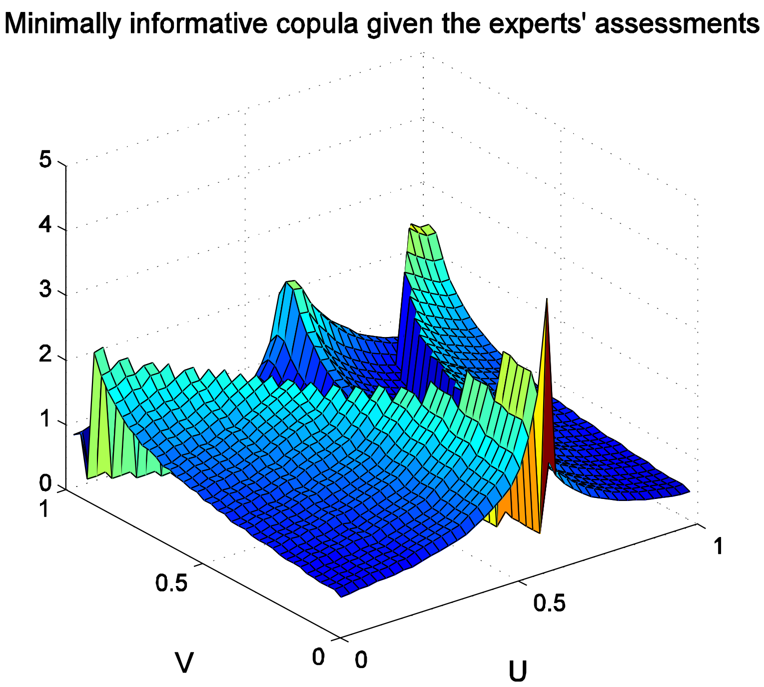
# Sequential – Step 2



# Sequential – Step 3



# Dependence elicitation:
## detailed quantitative assessments (1/2)

- We present a method that addresses the potential issues of *under-* and *overspecification* of detailed expert judgements

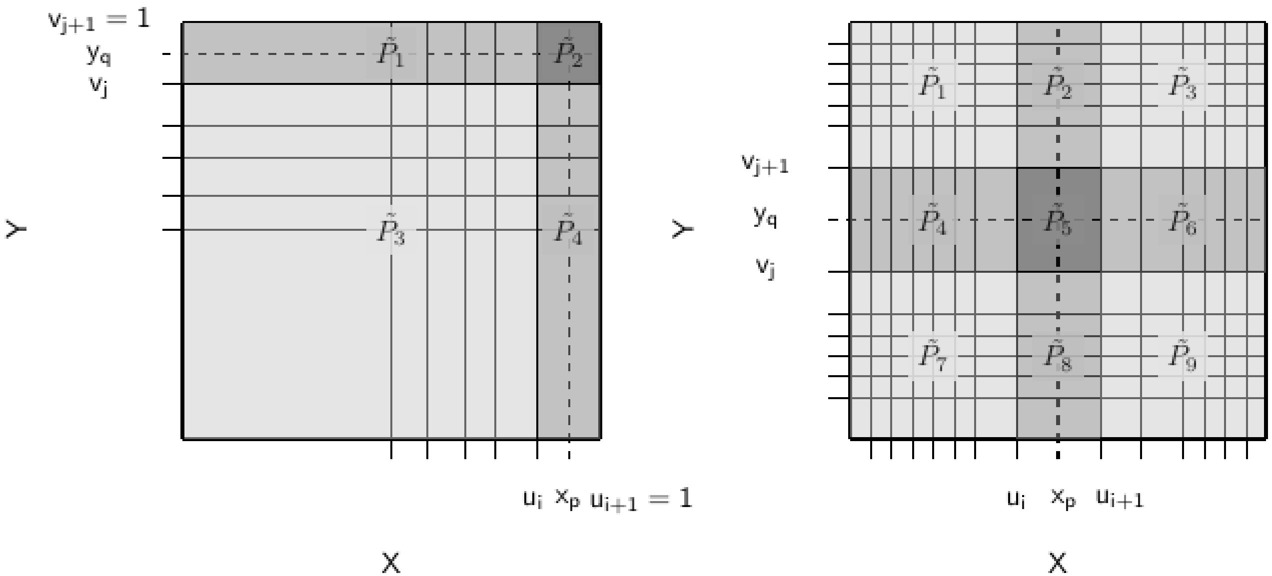
- For overspecification, an expert's assessments about related parts of a distribution are contradictory and infeasible; potentially occurring due to an increased cognitive complexity for experts when assessing a variety of detailed, related distribution features
- Underspecification means that we have not elicited enough information for modelling a unique distribution as various alternatives are compatible with the given (partial) information

- Proposed solution to overspecification: we only ever elicit single conditioning sets of low cognitive complexity and an algorithm providing the feasible ranges for any assessment is given
- Proposed solution to underspecification: assessed probability masses are modelled as minimally informative

Source: Werner C, Bedford T and Quigley J (2017) *The sequential refined partitioning method: addressing under- and overspecification of detailed expert judgement in probabilistic dependence modelling.* to be submitted

# Dependence elicitation:
## detailed quantitative assessments (2/2)

- A main contribution is the algorithm that provides the feasibility ranges for any assessment on the joint distribution
- Below two main examples are shown: (1) assessing the upper tail, (2) any additional judgement centrally



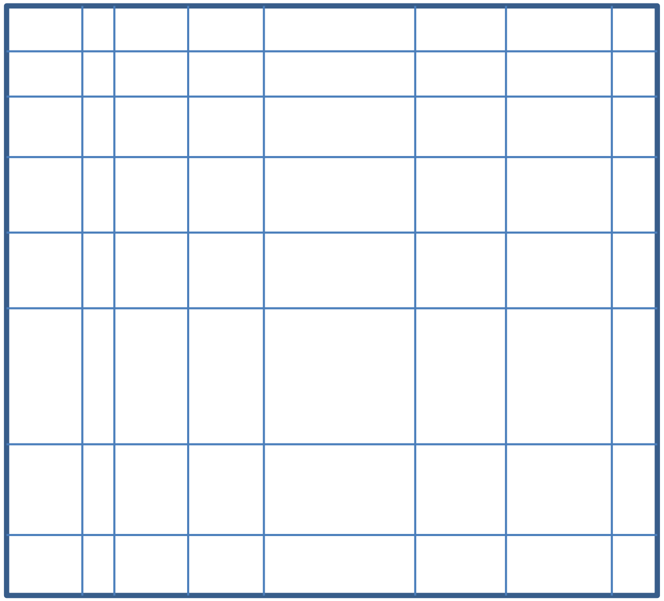

*Werner C, Bedford T and Quigley J (2017) *The sequential refined partitioning method: addressing under- and overspecification of detailed expert judgement in probabilistic dependence modelling.* to be submitted

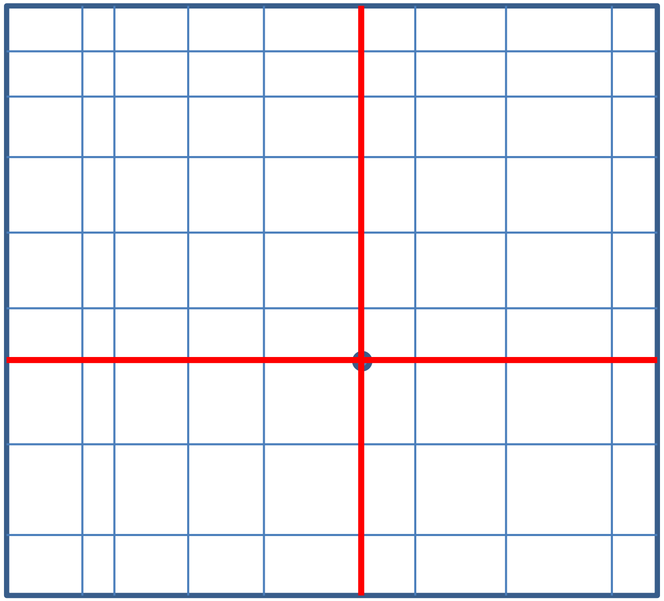
# Sequential elicitation

- Challenge of cognitive overload!
- Can we elicit dependence with simple questions of the form: $P(X > p|Y > q)$

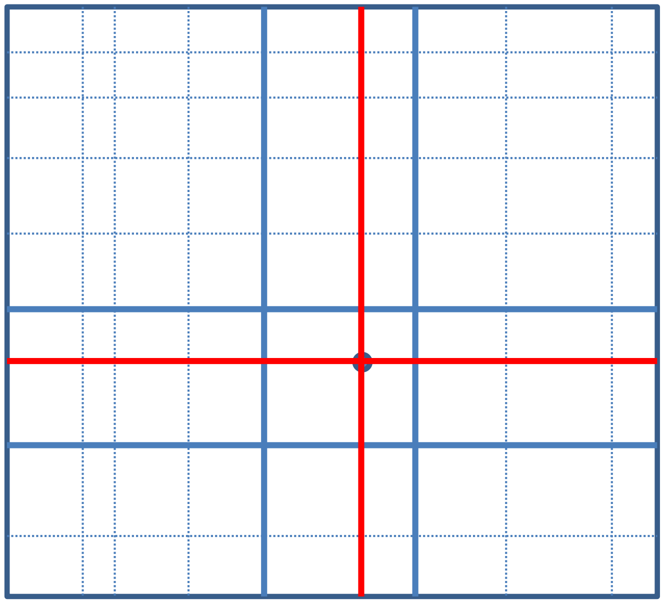
- Answer – yes, but there are lots of constraints
- Solution – can construct an LP problem with a limited number of constraints that provides exact bounds

# Approach – working on copula specification





# Approach



- 12 variables for probability on each block with red edge
- 2 column sum constraints
- 2 row sum constraints
- 4 block sum constraints

- Max/min the upper right red block probability to get constraints for expert

- Post judgement need to apportion probability into dotted blocks…

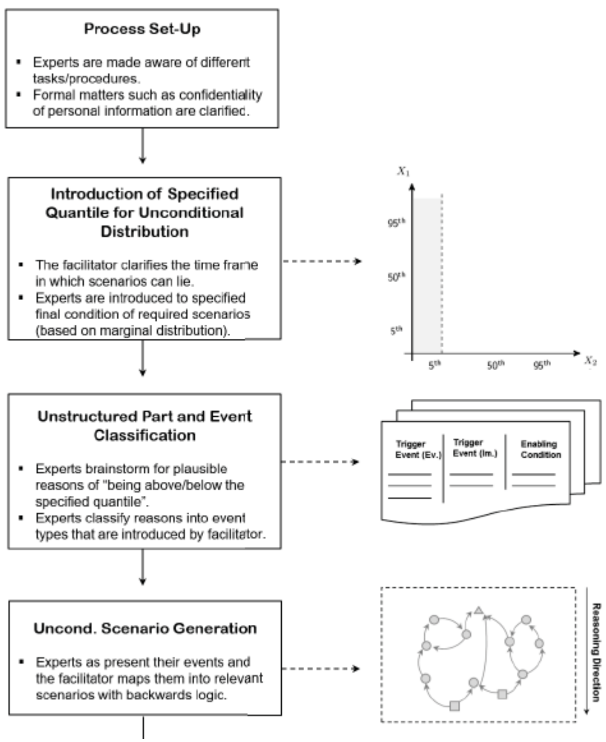
# Dependence elicitation:
## structuring experts' knowledge (1/4)

- We present a method for mapping *conditional scenarios*
- It allows experts to structure their knowledge on dependence relationships prior to a quantitative assessments
- The method can be used for a variety of dependence models/assessments and also for assessing tail dependencies
- Scenarios are defined for conditional dependence relationships
- There are indicative findings that the method allows for mitigating some common heuristics and biases that are prevalent in the assessment of conditional dependence parameters, such as (most prominently) confusion of joint and conditional probabilities and confusion of inverses, i.e. P(X|Y) with P(Y|X)

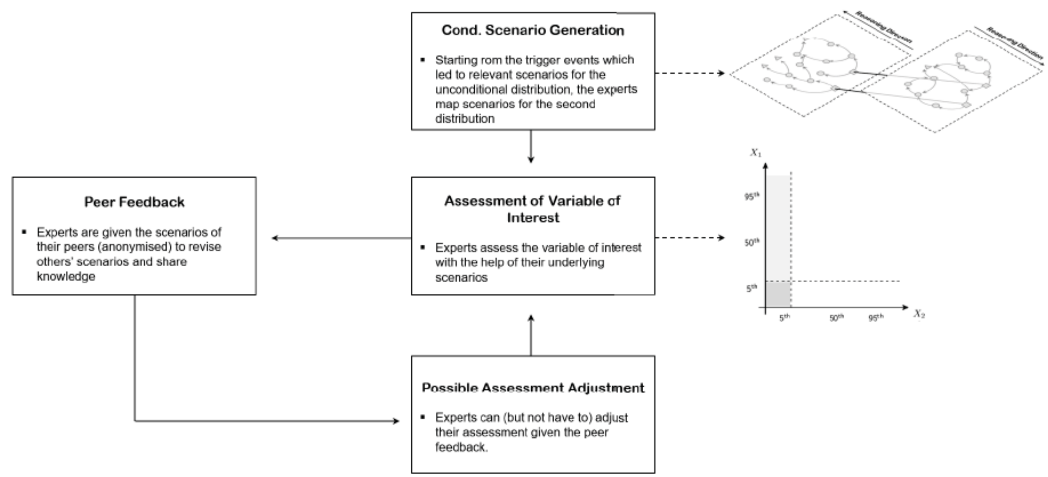
Source: Werner C, Bedford T and Quigley J (2017) *Mapping conditional scenarios for knowledge structuring in (tail) dependence elicitation.* Journal of the Operational Research Society, *under review*

# Dependence elicitation:
## structuring experts' knowledge (2/4)

Unconditional scenarios:




Conditional scenarios:



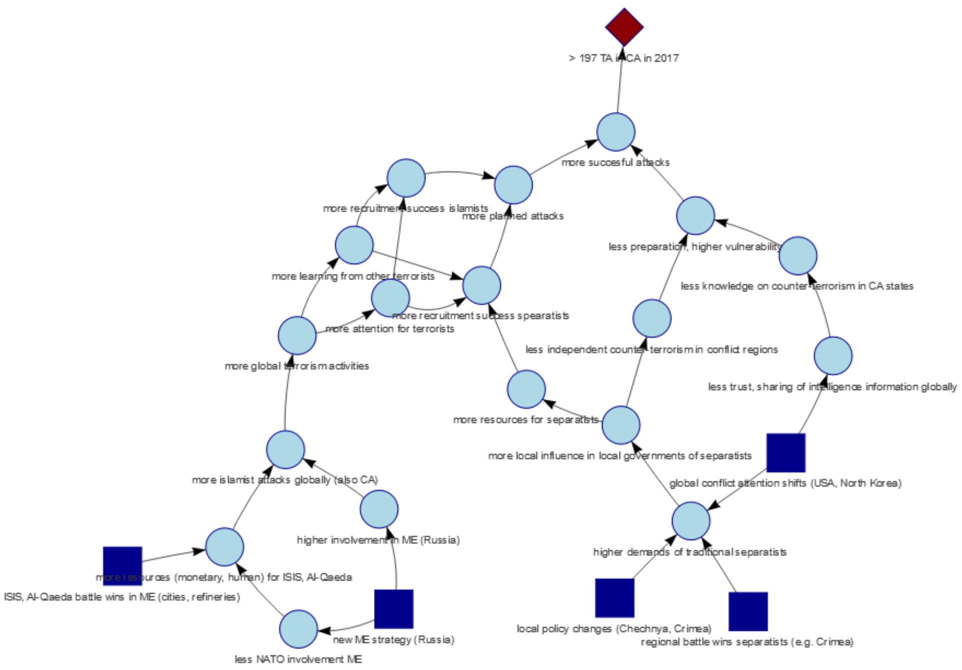

Source: Werner C, Bedford T and Quigley J (2017) *Mapping conditional scenarios for knowledge structuring in (tail) dependence elicitation.* Journal of the Operational Research Society, *under review*

# Dependence elicitation:
## structuring experts’ knowledge (3/4)

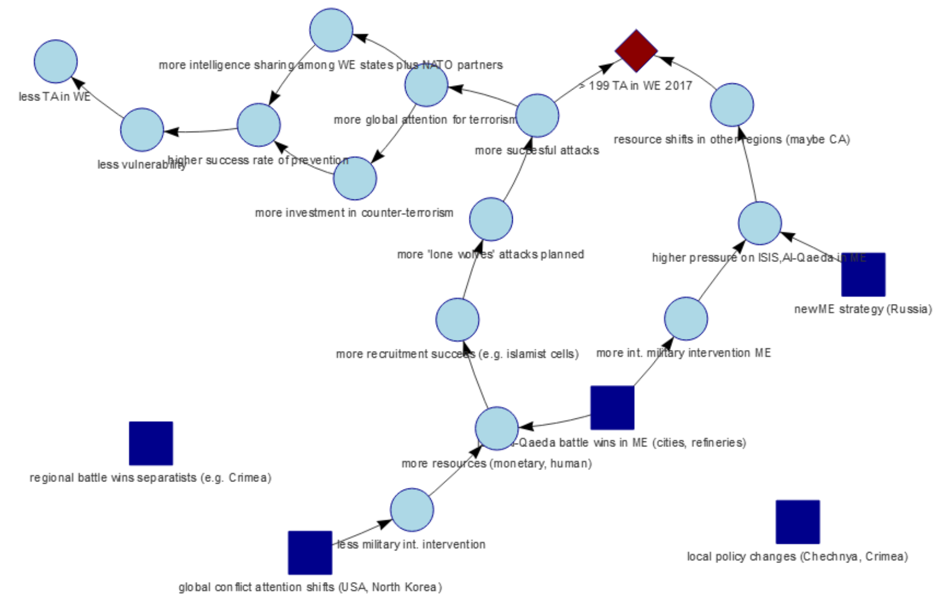
***Given that we observe more than 199 terrorist attacks in Central Asia (in 2017), what is the probability that we observe more than 225 terrorist attacks in Western Europe?***

Expert is professional catastrophe modeller/insurance underwriter on terrorism risk

**Unconditional Scenarios CA >95th**




**Conditional Scenarios WE>99th|CA >95th**



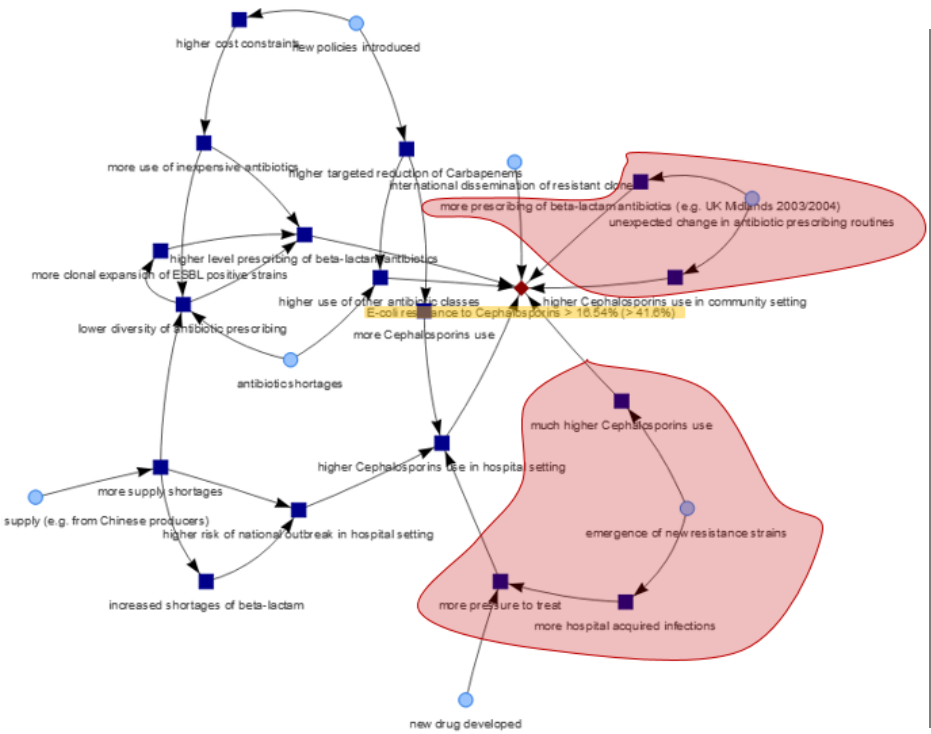

Source: Werner C, Bedford T and Quigley J (2017) *The sequential refined partitioning method: addressing under- and overspecification of detailed expert judgement in probabilistic dependence modelling. to be submitted*

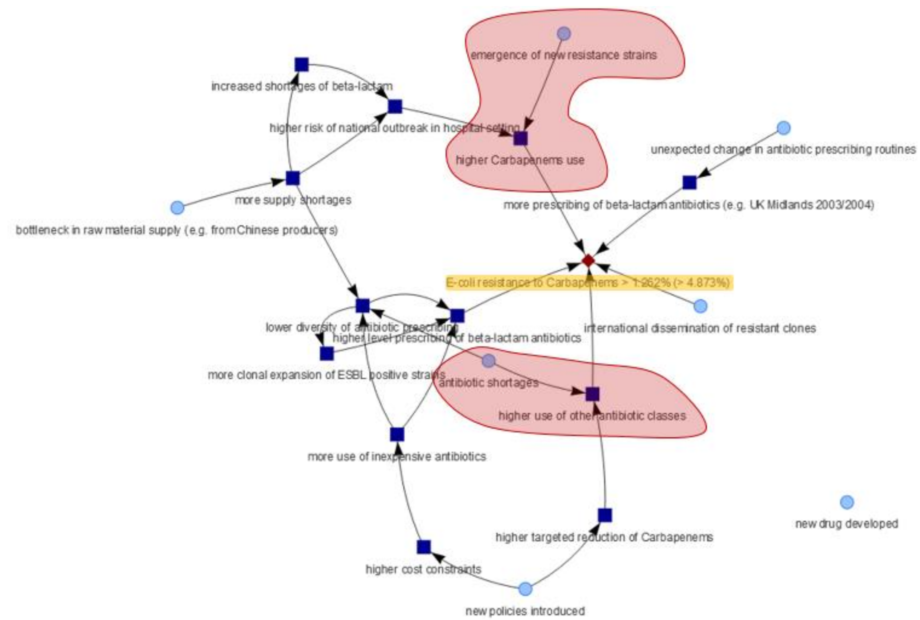
# Dependence elicitation:
## structuring experts’ knowledge (4/4)

University of Strathclyde Business School

***For the year 2021, given that in the UK the rate of Escherichia coli isolates resistant to third generation Cephalosporins is higher than 16.54 % [50th quantile] (41.6 % [95th quantile]), what is the probability that the rate resistance of Escherichia coli isolates to Carbapenems is higher than 1.262 % [50th quantile] (4.873 % [95th quantile])?***

5 experts who are medical researchers and practitioners in the area of antibacterial resistance in the UK




Source: Werner C, Bedford T, Colson A and Morton A (2017) *Risk assessment of future antibiotic resistance - eliciting and modelling probabilistic dependence between multivariate uncertainties of bug-drug combinations. to be submitted*

# Conclusions

- Dependency modelling a key, complicated, and (largely) under-researched area in risk and decision analysis
- To use in practice we have to take quantification step seriously
  - Little relevant data
  - Challenging for experts to consider
  - Methods needed that "lighted the load" for experts to reduce cognitive burden – complexity+time
  - Trade off between methods based on complexity and time impacts: Simple measures will need constraints, more complex measures could be constraint free
- Now a variety of elicitation processes available
- Use of maps for rationales seems to be welcomed by experts so far and useful tool to share information between experts.

*This work has benefited greatly from support of the COST network IS1304.*