SIMULATION BASED CAPITAL MODELS – TESTING, JUSTIFYING AND COMMUNICATING CHOICES

A report from the Life Aggregation and Simulation Techniques Working Party


ABSTRACT

The development of an economic capital model requires a vital decision to be made regarding how to aggregate capital requirements for the individual risk factors while taking into account diversification. Traditionally life insurers would adopt a Covariance-Correlation approach due to its simplicity and ease in communication to the stakeholders involved. The regulatory requirements of Solvency II have been one of the principal drivers leading to an increased use of more sophisticated aggregation techniques in economic capital models.

This paper briefly discusses the two broad types of aggregation techniques: closed-form (or deterministic) approaches and simulation based approaches, in addition to, the criteria that an insurer should consider when making their selection. It describes the Copula and Proxy Model method in detail, including the practical challenges in choosing and parameterising a copula, and challenges associated with fitting and validating a proxy model. In particular, the paper outlines how insurers could test, communicate and justify those choices by the use of some examples, to the various stakeholders involved.

Disclaimer

The views expressed in this publication are those of invited contributors and not necessarily those of the Institute and Faculty of Actuaries. The Institute and Faculty of Actuaries do not endorse any of the views stated, nor any claims or representations made in this publication and accept no responsibility or liability to any person for loss or damage suffered as a consequence of their placing reliance upon any view, claim or representation made in this publication. The information and expressions of opinion contained in this publication are not intended to be a comprehensive study, nor to provide actuarial advice or advice of any nature and should not be treated as a substitute for specific advice concerning individual situations. On no account may any part of this publication be reproduced without the written permission of the Institute and Faculty of Actuaries.
1 Introduction

1.1 Background

Some UK insurers have been using economic capital models to perform their own assessment of the capital required to support their risk exposures and to inform the management of those risks for a number of years.

The implementation on 31 December 2004 of realistic reporting for some UK with-profits firms and the Individual Capital Adequacy Standards (ICAS) framework introduced a risk based approach to the determination of regulatory capital requirements for UK life insurers, supplementing the previous factor based approach that applied previously under Solvency I.

One of the most fundamental choices to be made in development of any economic capital model is how to aggregate together the capital requirements for the individual risk factors and take account of the effects of diversification. Depending on the nature of the risk exposures and the level of granularity used in presentation, the reported effects of diversification can be very significant, typically amounting to a reduction of 40% to 60% of the sum of individual capital requirements.

A common approach to aggregation under the ICAS framework was to calculate standalone capital requirements for individual risk factors by applying stress tests – one for each risk factor. These individual capital requirements were then combined using a correlation matrix to determine an aggregate amount allowing for the effects of diversification. This result could then be subject to further adjustment to allow for some of the limitations associated with this approach (e.g. the implicit assumption that losses were linear functions of changes in risk factors) using the results of a scenario, calibrated using the individual stresses and correlations, to scale or replace the correlation matrix approach.

Under Solvency II, which will be implemented on 01 January 2016, UK life insurers must calculate their Solvency Capital Requirement (SCR) using a Standard Formula approach, or, subject to supervisory approval, use results produced by an Internal Model to substitute all or part of the Standard Formula calculation.

The Standard Formula approach of Solvency II uses a correlation matrix approach for the calculation of the SCR but, unlike under the ICAS framework, the correlation assumptions are prescribed.

In order for an Internal Model to be used in the calculation of the SCR, the Solvency II regulations require that the model meets certain minimum standards described in Articles 120 to 126 of the Framework Directive. These include standards relating to the statistical quality of the model, its calibration, a requirement for independent validation and the “use test” i.e. that the model plays an important role in informing decisions regarding the management of risk in the business. In particular, the calibration standards require that the SCR must be derived, where practicable, directly from the Probability Distribution Forecast generated by the Internal Model. That Probability Distribution Forecast is required to assign a probability of realisation to an exhaustive set of mutually exclusive events.

The Internal Model used by an undertaking must take into account the nature, scale and complexity of the risks to which it is exposed. Some UK life insurance undertakings have taken the view that, due to the nature of their risk exposures and the requirement for use of a “full Probability Distribution Forecast”, an aggregation approach based on a correlation matrix with scenario based refinements would not be adequate to meet the Internal Model standards of Solvency II. The regulatory requirements of Solvency II have been one of the principal factors which have led to an increase in the use of more sophisticated aggregation techniques in economic capital models.

These more sophisticated aggregation techniques are typically based around

1. Simulation of a large number of (pseudo) random scenarios from an assumed multivariate distribution of changes in risk factors. To achieve this, some insurers have adopted an approach based on copulas.

---

1 Solvency II Directive Article 13(38)
2 Pseudo-random Scenario: Scenario output satisfying many conditions expected from random numbers, although the generator operates in a deterministic way.
a “proxy model” to estimate the profits and losses which would arise in each simulated scenario. The proxy model approximates the profits and losses that would be produced by the “heavy” actuarial models in those scenarios, but can be evaluated much more rapidly.

These two components combine to produce a large number of simulated values of profits and losses from which the required measure of risk can be deduced.

The greater complexity of such techniques, the assumptions underlying them and their financial significance means that they are likely to come under greater scrutiny from the users of the models, such as senior management and the Boards of the undertakings. Where the models are to be used to determine regulatory capital requirements under Solvency II, they will also be subject to scrutiny by the supervisory authorities who will expect undertakings to be able to produce evidence that the approach meets all the relevant standards of Solvency II.

Some of the principal choices which are likely to come under particular scrutiny include:

- whether the model which describes the association or dependency between changes in risk factors and its calibration is appropriate – in particular, whether it makes adequate allowance for the association between extreme changes in the risk factors (“tail dependence”);
- whether the fitting error resulting from the use of the proxy model is material or appropriate adjustments are made – in particular, whether the proxy model makes adequate allowance for the effect of interactions between changes in risk factors on profit and loss.

The objective of the Life Aggregation and Simulations Techniques Working Party was to set out different techniques by which actuaries and insurers could assess and choose between the range of aggregation approaches available. In particular, the Working Party was asked to focus on how insurers could test, communicate and justify those choices to the various stakeholders involved.

The purpose of this paper is to provide UK life insurance actuaries with some examples of techniques which could be used to test and justify recommendations relating to the aggregation approach. We also aim to provide some examples of how those techniques may be communicated effectively to stakeholders. Whilst these techniques are more complex than those which have been common under the ICAS framework, the Working Party believes that the underlying concepts can be explained in a manner which is accessible to financially literate stakeholders without going into technical detail. We believe that graphical tools can be useful in explaining and justifying assumptions made and include some examples. We have tried to avoid discussion on the technical details and merits of specific techniques. However, we have included technical material or appropriate references where we believed this provides helpful context.

The Working Party understands that that the “Copula and Proxy Model” method is the most common of the more sophisticated approaches which some of the larger UK life insurers propose to use in their Internal Models, at least initially, under Solvency II. We therefore considered it appropriate to focus our attention on the challenges faced by actuaries when testing, justifying and communicating choices in relation to this approach.

1.2 Structure of this paper

Section 2.2 provides a high level overview of aggregation techniques, and describes in detail the criteria which they would be assessed against by a firm deciding which technique to adopt.

The remainder of the paper then focusses on the two main methods which are significantly distinct and which, to our knowledge, are in use by most UK life insurers:

- The Covariance-Correlation approach;
- The Copula and Proxy Model method.

The Covariance-Correlation approach is discussed in some detail in Section 3, although, as noted above, this is an established approach in the UK life industry and so we focus our discussion on the limitations of such an approach, and how these can be overcome.

The Copula and Proxy Model method is covered in Sections 4 and 5 of this paper:
Section 4 describes how copulas can be used as part of that methodology and the practical challenges in choosing and parameterising a copula. We discuss the application of statistical techniques used in other fields to assess the selection and parameterisation of a copula model, including a worked example. We also discuss how conditional probabilities (in the form of the coefficients of finite tail dependence) may be used to assess whether the calibration of a copula makes adequate allowance for tail dependence and introduce some graphical techniques which may assist in the selection and justification of assumptions.

Section 5 describes the challenges associated with fitting and validating a proxy model. We also discuss how the results of a proxy model can be communicated; and practical aspects of proxy modelling, such as allowance for reinsurance, tax or capital availability constraints.
2 Aggregation techniques and review criteria

2.1 Introduction

In this section, we provide an overview of the principal types of aggregation techniques available for the aggregation of capital requirements within an economic capital model. There are two broad types of techniques: closed-form (or deterministic) approaches and simulation based approaches; however some approaches exist which fall into neither of those categories. There are several variants of each type and we briefly describe these here. In subsequent sections of this paper, we discuss the most widely used variants in greater detail.

In the final part of this section we discuss how an insurer may approach the selection of an aggregation technique and criteria which should be taken into account when making such a selection.

2.2 Overview of Aggregation Techniques

2.2.1 Closed-form Techniques

Closed-form techniques provide an analytic formula for the capital requirement which is calibrated using a relatively small number of stress tests. The use of such a formula implicitly assumes a form of interpolation and extrapolation from the stress test results and relies on the following:

(i) Assumptions about the way in which the balance sheet responds to shocks in the risk factors;
(ii) Assumptions regarding the statistical distribution of changes in the risk factors.

To some extent, allowance can be made for these limitations through adjustments to the result of the analytic formula based on comparing the result of applying the formula to that produced by the full suite of “heavy” actuarial models in an appropriately chosen scenario.

The implicit use of a formula for capital results in an intertwining of the balance sheet response and risk factor changes. This means that if views on the distribution of an individual risk factor (and therefore a “1 in 200 year” shock to that risk factor) change, then one may need to re-run the heavy model suite to produce revised stresses in order to recalculate the capital requirements.

2.2.1.1 Correlation matrix approach

The correlation matrix approach was the approach adopted most commonly under the ICAS framework and is the method underpinning the Standard Formula of Solvency II. The 99.5\textsuperscript{th} percentile for each individual risk factor is calibrated, usually by choosing a distribution through a combination of fitting a model to data and judgement. The standalone capital requirement for that risk is then measured by modelling the corresponding stress using the heavy models. A correlation matrix, or “root mean square” approach, is then applied formulaically to calculate the overall capital requirement at 99.5\textsuperscript{th} percentile. (See Section 3 for further details.)

We discuss this approach in detail in Section 3 of this paper, where we note that it has a number of limitations. In particular:

- It also assumes that the balance sheet response to changes in individual risk factors is linear. For this reason, the approach is sometimes referred to as the “delta” approach.
- It assumes that changes in risk factors do not interact to compound or reduce losses - i.e. it assumes that the loss arising from a change in two more risk factors is equal to the sum of the losses resulting from changes in each risk factor individually. (This is sometimes referred to as “separability”.)
- Together, these two assumptions imply that the Value Response Function (i.e. the formula that expresses losses in terms of changes in risk factors) is linear.
- The closed-form formula is precise if the distribution of the changes in risk factors takes a specific form known as an “elliptic” distribution (e.g. a multivariate Normal).

Depending on the nature, scale and complexity of the business, these limitations may make it more challenging for such an approach to meet the relevant Statistical Quality Standards.
2.2.1.2 Delta/Gamma approach

The Delta/Gamma approach is an extension of the pure correlation matrix approach where, instead of assuming a linear Value Response Function, the formula now includes quadratic terms, including interactions between pairs of risk factors.

The method is calibrated by applying two stress tests in each of the individual risk factors and one joint stress test for all combinations of two risk factors.

Under the assumption that the changes in risk factors follow a multivariate Normal distribution, a closed-form solution is provided for the resulting Value at Risk measure.

This approach mitigates some of the limitations of the simpler “delta” approach at the expense of requiring more runs of the “heavy” models.

The Working Party is unaware of the Delta/Gamma approach having been used in practice by any UK life insurance undertaking in its economic capital model. However there are Swiss firms using this approach which is standard under the Swiss Solvency Test.\(^3\)

2.2.2 Simulation Techniques

2.2.2.1 Introduction

In the previous section, we noted several limitations of closed-form techniques to capital aggregation:

- The intertwining of the joint distribution of risk factor changes with the response of the balance sheet to those changes.
- The assumption of a linear balance sheet response to shocks in individual risk factors.
- The assumption of no interaction between simultaneous shocks in two or more risk factors to compound or mitigate their individual effects.
- Strong assumptions on the form of the joint distribution of risk factor changes.

In particular, we noted that the correlation matrix approach is accurate if the joint distribution of risk factor changes is elliptic (e.g. Normal) and the Value Response Function is a linear function of the risk factors. However, this need no longer be the case if the distribution of risk factor changes is not elliptic or the Value Response Function is non-linear.

A simulation based approach to aggregation can address these limitations. By a simulation based approach, we mean an approach to capital aggregation that consists of two components:

(i) A simulation based Risk Factor Model

This produces a large number of simulated scenarios corresponding to changes in the risk factors allowing for the relationship (or “dependency structure”) between changes in the individual risk factors.

(ii) Value Response Model

This explains how the balance sheet responds to changes in the risk factors.

By evaluating the Value Response Model at each of the simulations produced by the Risk Factor model, a distribution of losses is produced. The desired measure of risk is then extracted from that distribution (e.g. if a Value at Risk measure is being used, the relevant percentile can be extracted).

The “de-coupling” of the Risk Factor Model and Value Response Model brings some advantages.

The Risk Factor Model permits flexibility in the choice of model for individual risk factors and the dependency between them.

The Value Response Model can allow for non-linear response to shocks in individual risk factors and the compounding or offsetting effect on losses of interactions between risk factors.

\(^3\) Delta-Gamma-Verfahren als StandardRisikomodell für Lebensversicherer (See [ref. DGV])
There are several variants of simulation based approaches which depend on the combination of choices made for the Risk Factor Model and Value Response Model. We describe these briefly below. The simulation based approach which has been most commonly adopted by undertakings intending to use an Internal Model to calculate their SCR under Solvency II combines the copula based approach for the Risk Factor Model and the Proxy Model approach for the Value Response Model. These models are discussed in more detail in Sections 4 and 5.

In addition a modular approach is possible, particularly for Partial Internal Model undertakings. Here the copula based approach and Proxy Model approach are used for a subset of risk factors. These are then combined with the other risk factors using a Covariance matrix approach. This may address the non-separability between some risk factors and the non-elliptical distributions of some risk factors, without introducing complexity where it is unnecessary.

2.2.3 Risk Factor Models

2.2.3.1 Copula models

Under this technique, a combination of historical data and expert judgement are used to identify a suitable statistical distribution for each material, quantifiable risk factor. A copula is chosen to model the dependencies between changes in those risk factors. A large number of simulated changes in each individual risk factor is generated and “glued together” using the copula to generate a set of simulated scenarios which reflect the assumed dependencies.

2.2.3.2 Structural models

Another way of coming up with a distribution of changes in risk factors over a given time horizon is use a so-called “structural model”\textsuperscript{4}. These are typically based on time-series forecasting models which generate sample paths of changes in risk factors over small time steps, making assumptions about the changes observed in previous time steps combine with random shocks to generate changes in the current time step. The joint distribution of changes in the risk factors over a given time horizon can be extracted from the corresponding point of the time series forecast.

For example, one approach would be to use an appropriate calibrated Real-World Economic Scenario Generator and augment it by inclusion of additional factors describing non-economic risks and assumptions about the inter-relationships involving those risk factors.

Such an approach can have the advantage of a clearer “cause and effect” relationships between changes in successive time periods, resulting in a coherent and more realistic dependency structure.

The main challenges relate to transparency and the level of expert knowledge required to maintain such a model and avoid it becoming a “black box”.

2.2.3.3 Causal models

The third approach looks to the underlying causes that drive the joint behaviour of risk factors. These causal models are typically produced using Bayesian Network techniques\textsuperscript{5}. The Working Party is only aware of causal models in use in the UK Life industry for the purpose of modelling Operational risk capital. One advantage of the Bayesian Networks technique is that it facilitates the allocation of capital back to the underlying causes.

2.2.4 Value Response Models

2.2.4.1 Heavy Actuarial Models

One option to calculate capital requirements is, as has been done in the past, to calculate results using the full heavy actuarial models. For some with-profits products featuring path-dependent guarantees, a stochastic model

\textsuperscript{4}A stochastic investment model for actuarial use. D. Wilkie. 1984. See [ref. WWW]

\textsuperscript{5}Risk Factor Portfolio Management. Milliman. 2015. See [ref. MWW]
is required to value the business\(^6\), and the result is the average over a large number of equally likely scenarios. This creates the “nested stochastic” problem, where, in each simulation, a stochastic model must be run to value the business in that particular scenario. Due to current computer run times, nested stochastic calculations can make it infeasible to perform a large number of runs within a reporting period. In practice firms will want run-times to be much shorter than this so that they can evaluate what-if business decisions and satisfy the use test. Further, with the large number of simulations from the Risk Factor Model that are required to ensure that no material estimation error exists in the calculation of economic capital requirements, it may not even be feasible to run the heavy deterministic models for each simulation. However, one option is for firms to invest in computing power to enable them to perform deterministic and nested stochastic evaluations more quickly. With development in technologies growing at the fast pace as we saw in the past two to three decades, it is possible that nested stochastic simulations will become technologically feasible. While the exact technology will hardly be a timely way-out for the 2016 “due date” of Solvency II in the UK, “brute-force” nested stochastic capital calculations is an alternative that may become feasible in the future.

2.2.4.2 Proxy Models

A proxy model is a function of changes in the risk factors which provides a suitable approximation to the results of the heavy models. Its use aims to allow for the practical difficulties with running the heavy model in each simulated scenario described in the previous section.

Typically, an undertaking will construct the proxy model by splitting its business up into a number of portfolios of assets and liabilities and using suitably chosen functions such as polynomials to approximate the results of the heavy models for those portfolios. This technique is sometimes known as “curve fitting”. For example, an undertaking may fit a polynomial to the change in value of assets backing its annuity book under a shift in the yield curve.

The proxy models are usually calibrated by fitting the chosen functions to the results of heavy models evaluated at a set of calibration scenarios for each portfolio.

2.2.4.3 Least Square Monte-Carlo

Least-Squares Monte-Carlo (LSMC) is a special form of curve fitting which can help to address the nested Monte-Carlo problem described above. In LSMC, a large number of fitting (or “outer”) scenarios are chosen but each fitting point uses a small number of “inner” runs of the stochastic model to value the business. The estimated liability values are therefore generally less accurate than the heavy model result in each “outer” scenario. Regression techniques are then used to fit an appropriate proxy function. The resulting proxy function smooths out the estimation errors in the outer scenarios.

2.2.4.4 Replicating portfolio

This technique involves optimisation algorithms to find a notional portfolio of financial instruments that best mimic the behaviours of the insurance liability and whose value can be expressed in a form which can be rapidly calculated (e.g. by an analytical pricing formula or highly accurate closed form approximation). This involves identifying a portfolio of such instruments that has similar sensitivities to those of the relevant insurance liabilities against a range of mainly economic risk factors.

The main drawback of this technique is that some with-profits insurance liabilities are usually of highly complex nature, involving path dependencies and dependencies between each other. This can make the task of finding a notional portfolio that accurately describes the behaviours of a portfolio of with-profits liabilities very challenging.

\(^6\) Typically these models use model points (grouped cells via compression or ad-hoc) rather than individual policy records to further cut down runtimes.
2.2.5 Other Aggregation Techniques

2.2.5.1 Risk geographies

This technique uses the geometry underlying an elliptical distribution to provide a measure of the severity or likelihood of an individual scenario. It identifies a scenario which is in some sense “most likely” and which gives rise to losses at the level of the economic capital requirement. It does this by a process of iteration. The Value Response Function is approximated by a suitable proxy model, the resulting capital requirements estimated and the most likely scenario identified. The proxy model is then refined by recalibration to sensitivities at the most likely scenario, the capital requirements re-estimated and the most likely scenario recalibrated. The process is iterated until the capital requirements and most likely scenario converge to a suitable degree of accuracy.

To our knowledge, this technique has so far not been widely used within the UK life insurance industry.

2.3 Selection of an aggregation technique

The choice of aggregation technique is fundamental to the way in which economic capital is modelled and the technology and processes involved. Each technique has its strengths and limitations and the choice may not be straightforward. In forming their recommendations to Boards and senior management, actuaries will therefore need to give careful consideration to a range of factors and criteria. One way of approaching would be as follows:

(i) Agree to a short list of aggregation techniques and a set of criteria to be met with the decision makers.
(ii) Evaluate each of the techniques against those criteria. The results of this evaluation could be presented in tabular form.
(iii) Formulate a recommendation based on the results of the evaluation.

We set out below some of the criteria and factors which may be considered.

2.3.1 Nature, scale and complexity of the business

The extent to which a specific approach is capable of reflecting the nature and complexity of the business is an important factor.

For example, the following may indicate the presence of non-linear balance responses:

- The existence of a large block of with-profits business with complex assumed management actions – the guarantees on this business are embedded derivatives whose behaviour is complicated by management actions which may have specific triggers and be subject to caps or floors.
- Non-proportional reinsurance arrangements such as excess of loss treaties
- Hedging strategies

There may also be arrangements which restrict the ability of capital held in one part of the business to absorb losses arising in another part. Such features may reduce the effects of diversification. For example:

- A with-profits fund typically results in constraints on the use of any estate of that fund
- Ring fenced funds, including the matching adjustment portfolio
- In the case of branches subject to regulation by some overseas authorities, there may be a requirement to hold a minimum value of assets locally.

These features are more difficult to model using a correlation matrix based approach and may suggest use of a more sophisticated technique.

However, the scale of the risks and the principal of proportionality are also important. It may not be proportionate to use a sophisticated and costly model if complex features relate only to a relatively minor line of business. For a small company with relatively simple risk exposures which is part of a larger group, a relatively simple approach may be adequate.
2.3.2 Use in managing the business

Is the model to be used for calculating regulatory capital requirements under Solvency II? If so, then it must meet all the relevant standards (see below).

**Regular solvency monitoring** – if the model is to be used frequently for monitoring the economic capital position, then it must be capable of being re-run either in full or with suitable approximations at the required frequency. Simulation techniques which separate the Value Response Model from the Risk Factor Model can accommodate this by “rebasin” the Value Response Model to reflect changed conditions, making appropriate allowance for the run-off of in-force business and writing of new business by rescaling of loss functions. For correlation matrix based techniques, adjustments may be made to standalone capital requirements to reflect changes in conditions and the business. These may be less transparent, but may be fit for purpose, depending on the nature of the business.

**Stress and scenario testing** – techniques based on proxy models may be more useful if the company makes frequent use of stress and scenario testing. Provided the arguments of the loss functions have been appropriately chosen, the loss functions can be evaluated in the chosen scenarios, so avoid full runs of the heavy models.

**Capital management** – simulation based techniques provide the ability to “slice and dice” the simulated loss output in various ways which can generate a richer set of Management Information. For example, it is more straightforward to see the “shape” of the distribution of losses by looking at different percentiles of the Probability Distribution Forecast. Capital requirements for sub-portfolios can be identified separately if the loss functions are specified at an appropriate degree of granularity. It is possible to establish the contribution of individual sub-portfolios to the overall capital requirement by averaging out losses in some way (e.g. kernel smoothing) in a window of simulations around the simulation that bites at entity level. This can help produce a “heat map” of risk hot-spots after allowing for the effects of diversification, giving management more information to form their view on where to focus risk mitigation or risk taking activity.

**“What-if?” scenarios** – the assessment of the implications of a change in the business (e.g. the introduction of a hedge or a change in management actions) can be more difficult if using a correlation matrix approach as it may require a re-run of several stress tests. One approximate method is to test the effect in one of the scenarios used to refine the results from the correlation matrix. However, if the change is expected to result in a significant shift in this scenario, then there will be greater uncertainty over the result. For some types of change, the greater flexibility of a proxy model approach may be more useful. For example, if a change relates to only one portfolio of risk, the loss functions for that portfolio could be adjusted to reflect the change and the simulation model re-run.

2.3.3 Transparency and complexity

Boards and senior management will want to be satisfied that the model is fit for the purposes for which it is used. They will therefore want to understand the general structure of the model, the significant judgements made and their financial impact as well as the limitations of the model, including circumstances in which the model is less reliable. In general, a model should be no more complex than can be justified by the purposes for which it is to serve.

Simulation based aggregation models are significantly more complex and involve more explicit assumptions and judgements than a correlation matrix based approach. Actuaries recommending a model will need to balance this complexity against the use to which the model will be put and the nature, scale and complexity of the business.

However, in making some of the assumptions (e.g. the distributions of individual risk factors) more explicit, simulation based models can be more transparent and lead to increased engagement of senior management in selecting assumptions. The richer information produced by such models can also lead to greater discussion and enhance understanding about the nature of the risks run by the business. For example, the simulations around the “biting scenario” provide concrete examples which can help focus discussions and provide a valuable tool for validating the output of the model. Examination of the loss functions can assist in the understanding of how the balance sheet responds to risks and how different risk factors interact.
2.3.4 Regulatory requirements

If the economic capital model is to be used for regulatory reporting (e.g. to calculate the SCR under Solvency II), then there are additional considerations to take into account. For example, Solvency II sets out a number of requirements in the form of the Framework Directive, Delegated Acts and EIOPA Guidelines. Article 121 of the Framework Directive sets out a number of Statistical Quality Standards explicitly states that no particular method of calculating the Probability Distribution Forecast shall be prescribed. However, the Directive and other regulatory texts set out a number of general requirements. We set out some of those most relevant to aggregation below. The reader is referred to regulatory texts for more details.

2.3.4.1 Richness of the Probability Distribution Forecast

Article 101 of the Framework Directive requires that the SCR be equivalent to the 99.5th percentile of reduction in Basic Own Funds over a one year time horizon. Article 13(38) requires that the underlying Probability Distribution Forecast “assigns to an exhaustive set of mutually exclusive future events a probability of realisation”. This could initially be interpreted as requiring a distribution which is calibrated at a dense set of scenarios, which may rule out correlation matrix based approaches which use a limited number of stress tests and make assumptions about the form of the underlying distribution to interpolate and extrapolate the Probability Distribution Forecast from these.

However, Article 228 of the Delegated Acts states that “the mutually exclusive future events … shall contain a sufficient number of events to reflect the risk profile of the undertaking”. The EIOPA Guidelines on The Use of the Internal Model state that “the insurance or reinsurance undertaking should choose techniques that generate a Probability Distribution Forecast that is rich enough to capture all relevant characteristics of its risk profile”.

It is our interpretation that the Solvency II regulations therefore do not require using a simulation based approach to generate a “full” Probability Distribution Forecast and do not exclude the use of a correlation matrix based approach, provided that the undertaking is able to demonstrate that the resulting Probability Distribution Forecast appropriately reflects its risk profile.

Undertakings will therefore need to reflect the nature of the risks to which they are exposed and the way in which they would expect their balance sheet to respond to those risks and how these compare to assumptions underlying the models under consideration.

It may be more challenging for undertakings that have significant exposures to risks where there is evidence to support departure from a Normal distribution or where it is expected that the balance sheet would respond in a non-linear manner to shocks to justify the use of a correlation matrix based approach.

2.3.4.2 Adequacy of system for measuring effects of diversification

The Statistical Quality Standards require that the undertaking must be able to satisfy the supervisory authorities that the system used to measure the effects of diversification is adequate. This is expanded upon by Article 234 of the Delegated Acts which, inter alia, requires that the system for measuring diversification takes into account any non-linear dependence and any lack of diversification under extreme scenarios, as well as any, restrictions of diversification which arise from the existence of a ring-fenced fund or matching adjustment portfolio.

The aggregation technique used must therefore be able to reflect non-linear balance sheet responses, and take account of tail dependence and constraints on fungibility of Own Funds.

2.3.4.3 Justification to supervisory authorities

The Statistical Quality Standards require that firms are able to justify the assumptions underlying their internal model to the supervisory authorities. It is therefore important that the rationale for the selection of a particular aggregation approach and the justification for the related assumptions is adequately robust.

2.3.4.4 Reporting timescales

Reporting timescales under Solvency II are challenging. For example, annual returns for solo undertakings will ultimately have to be published within 14 weeks of a firm’s reporting year end. These timescales have significant implication for the design and operation of an Internal Model, including the choice of aggregation
More complex techniques which rely on large numbers of heavy model runs or which involve several iterations of modelling may make reporting within the required timescales more difficult, unless this can be overcome by use of other techniques such as use of "roll-forwards".

2.3.5 Industry practice

Choice of a technique which is less popular commonly leads to overheads. It may be more difficult to justify the model to stakeholders such as Boards and supervisory authorities and more difficult to compare the underlying assumptions with other companies and understand any differences in results. These differences lead to additional costs. Additional costs also arise in training new staff if turnover leads to external recruitment.

2.3.6 Operational processes

Greater model complexity can lead to more cumbersome processes and can increase operational risk. Knowledge of the model needs to be maintained; staff, including Board and senior management must be educated about the model to an appropriate level; assumptions need to be maintained and appropriately updated, the model may require new IT platforms to operate on; and reporting processes may need to be revised, in particular where the model is to be used for external reporting purposes.

2.3.7 Cost

The cost of operating and maintaining a model may be a factor. There will be costs in the initial development and implementation of a technique but also in relation to its ongoing maintenance.

2.4 Conclusion

The selection of an aggregation technique is one of the fundamental choices in the design of any economic capital model. All aggregation techniques have strengths and limitations and it is important that these are taken into account when making the selection to ensure that the model is fit for the purposes for which it will be used.

We refer the reader to a recent paper by the Model Risk Working Party that considers some of these issues in further detail (See [ref. MRWP]). Further, the reader may wish to consult the recent paper on Expert Judgement by the Solvency & Capital Management Working Party (See [ref. SCMWP]) ; as well as Solvency II Delegated Acts; and the Technical Actuarial Standards, in particular TAS M.
3 Covariance-correlation approach

3.1 Introduction

Covariance-correlation is an analytical framework for aggregating capital that has been used widely by UK firms since the introduction of the ICAS regime in 2004. Under this method univariate stresses are run on the balance sheet and capital requirements for each risk are calculated at the 99.5th percentile. These univariate capital requirements are then aggregated through a correlation matrix.

Note that correlations between risk factors are not the same as correlations between losses (capital requirements) as life insurance losses do not move linearly with the risk factors. Where there is ambiguity, we will refer to the former as risk-correlations and to the latter as capital-correlations. Variations of the technique have been used for a while in the UK under the ICA regime. It’s one of the simplest methods and the same principle is used to calculate the SCR using the Standard Formula method in Solvency II.

In the context of internal models, it could be used to allow for risks not captured under the Standard Formula or risks captured under the Standard Formula, but where the calibration is not appropriate in reflecting the nature of each specific risk the company is actually taking. Under Solvency II, the Standard Formula calculation could be augmented with additional company specific risks, to create a ‘Partial Internal Model’. The Solvency II guidelines include the option to expand the Standard Formula modules and correlation matrices to include additional these additional risks. (Alternatively, Standard Formula risk calibrations and correlations can be incorporated into an Internal Model framework).

A possible formula for this approach is:

\[ SCR_{Total} = \sqrt{\sum \rho_{ij} SCR_i SCR_j} \]

where \( \rho_{ij} \) denotes the correlation between risk factor \( i \) and \( j \) and \( SCR_i \) denotes the standalone required capital under the univariate stresses \( i \).

3.2 Motivation for the approach

The covariance-correlation approach might be deemed to be an appropriate method where risks follow an elliptic distribution (such as a multivariate normal) and there is no significant non-linearity or other interactions between risks; it might also be the preferred approach for some insurers due to its simplicity. The approach might be deemed to be more transparent by some after allowing for the assumptions implied in the method; and it is easier to make like-for-like comparisons, for benchmarking against peers. The method is also useful as a validation check against the copula+proxy model method (discussed in sections 4 and 5) and for producing sensitivities.

For some insurers, the copula+proxy model development costs may outweigh the benefits and a firm may prefer to use the simpler Covariance approach, either as is, or by making appropriate refinements to reflect features specific to its risk profile (e.g. add-ons based on scenario analysis, or use more prudent correlation factors). This is more likely to be the case for smaller companies, which are not deemed to pose a systemic risk to a country’s economy. However, costs of the copula+proxy model approach may fall over time as it is being used more widely.

However, as with most simplified approaches, there are a number of more material limitations, and we will discuss these below.

3.3 Limitations of the approach

In its purest form, the method implies that the multivariate risk factor distribution is elliptical (an example would be normally distributed risk factors correlated through a Normal (Gaussian) copula) and that the relationship between risk factors and univariate capital losses is linear. However these assumptions are not usually borne out in practice, and would be difficult to justify under the Solvency II Statistical Quality
Standards. We should note that although we describe next these assumptions, we have not tried to quantify their effect. A study has quantified some of the deviations of the approach see [ref. SSS].

It is not certain that it is appropriate to model univariate losses using a normal distribution; for example, historical data suggests that equity returns are heavy-tailed. Equity level risk factors are sometimes modelled through fat-tailed distribution, for example lognormal or hyperbolic.

For some insurance contracts and particular risks the linear relationship might hold (e.g. equity risk and linked liabilities), but there will be examples where this relationship will break down. This can be as a result of guarantees of some kind (e.g. minimum benefits) or stop-loss reinsurance contracts, or in the case of interest rates where valuations exhibit convexity. Management actions can be another example, where such actions are only triggered in extreme events.

The method also assumes that the risks are separable, i.e. the univariate losses from each risk plus the correlation matrix are sufficient to determine the aggregate loss. This is not always the case, as there will be additional effects from considering the two risks together, which can be material. Examples can be the relationship between interest rates and longevity on an annuity contract and levels of equity returns and interest rates on the “Value in Force” of a unit linked contract. The impact from non-separability (usually also mentioned as “non-linearity”, of which it is simply a particular instance) can vary between different pairs of risks but it will be significant at least for some. There is also the risk of double-counting effects of management actions that may take place in different scenarios, but they are the same actions (for example reducing bonus rates may be an action taken when either equities fall or (say) credit spreads risk, but in a combined scenario where both events take place, the effects of the action should only be accounted once). Double-counting may also arise with other risk mitigating techniques in place. Also where there are multiple ring-fenced funds, non-linearity may affect losses in opposite directions between different funds.

As a result of the above, risk-correlations based on observable risk factors (e.g. equity prices, mortality rates etc.) which are non-normally distributed will only by chance be similar to the capital-correlations that correspond to these risk factors.

Overall, there are a number of limitations within the method, but it can still produce more valid results, if certain efforts are used to counteract the limitations. We are looking at these next.

3.4 Compensating for limitations of the method

A number of techniques can be employed to compensate for the limitations of the approach. We describe these next. The techniques are described in broad terms and they need to be appropriately tailored to reflect the circumstances of a particular insurer.

3.4.1 Non-linearity

Combining 99.5 percentiles through a covariance matrix may result in a more (or less) penal percentile in the aggregate loss distribution than the 99.5th when non-linearity is present. A couple of approaches attempt to adjust for this.

The “big bang” approach compares the results of (i) a scenario which combines changes in all risk factors at a 99.5th percentile level with (ii) the sum of the standalone capital requirements for those factors at a 99.5th percentile level. The ratio of (i) to (ii) is a measure of non-linearity which is then applied to scale the results of the correlation matrix.

The “big bang” scenario does not recognise the effects of diversification between risks and so is likely to result in a combined scenario which is more extreme than the scenarios giving rise to “1 in 200 year” losses.

The “medium bang” approach makes an approximate allowance for diversification by using a lower percentile (93rd say) selected by judgement. The scaling factor is used to adjust the result of the correlation matrix in the same way.

A more refined approach (sometimes referred to as the “magic” or “Killer” scenario) uses the Euler approach to allocation of the effects of diversification. This produces a scenario consisting of a combination of changes in risk factors, but where the percentile at which each risk factor appears depends on its contribution to the correlation matrix result. If the assumptions underlying the correlation matrix approach are correct, this
scenario is in some sense the “most likely” scenario that gives rise to a loss equal to the aggregate capital requirement.

3.4.2 Other features

Allowance for management actions can be made by appropriate alterations to the heavy models, either explicitly or through scaling. Scenarios where many risk factors are simultaneously stressed might be used.

Fungibility, transferability, various types of reinsurance and guarantees can also be allowed for approximately, through scenario testing or by observing features and the behaviour of losses against changes of risk factors for the business modelled (restrictions in fund movements, monotonicity and local minima/maxima of losses, “cliffs” due to the presence of excess-loss reinsurance or guarantees to policyholders, management actions to be taken in extreme scenarios) and making appropriate adjustments. Adjustments can be applied to either univariate losses or to the aggregate loss amount that represents the 99.5th percentile.

However, allowing for such features is most likely to result to some degree in approximations and judgement in identifying the 99.5th percentile of losses, given that some of the features above are only relevant for certain events. It will also be very difficult to keep a clear view of what is the defining 99.5th percentile, as refining the model itself re-defines the 99.5th percentile.

Further capital add-ons may be used to allow for any features not properly captured through the above. If the capital add-on is self-imposed, this introduces a further layer of expert judgement. Capital add-ons could also be imposed by the regulator, if the regulator deems there is an insufficient level of self-imposed capital add-ons.

3.5 Variations of the main technique

Variations of the covariance-correlation method can allow (with varying levels of accuracy) for

- A multi-layered approach (see section 3.5.1)
- Non-linearity and non-separability (e.g. the delta-gamma approach, see section 3.5.2)
- Tail dependency (see section 3.6.1)

These are discussed below.

3.5.1 Multiple layers of aggregation

At its simplest form, the covariance-correlation method has a single layer of aggregation, whereby all univariate capital losses are linked through a single capital-correlation matrix (see Left diagram of Figure 3-1). A common variation (also used in the Standard Formula, under Solvency II) is to use two or more layers of correlation. The risks within the innermost layer usually share some common characteristics, (for example level and trend of the mortality risk, or sub-risks of market risk, such as equity, credit spread, interest rate, exchange rates, inflation).

The method can be extended to accommodate further layers (for example drivers within interest rates, or credit spreads that reflect different credit ratings); an illustrative structure is the right diagram of Figure 3-1. The method would then be to first compute univariate losses at the innermost layers and then aggregate them towards the outermost layer by using the appropriate correlation matrices.
Using the multi-layer approach may appear as a more intuitive way to derive correlations and reduces the dimensionality problem of an oversized correlation matrix. It can be a difficult, and arguably spurious, exercise to calibrate the correlation between two very granular risk factors, for example annuitant mortality trend and property prices. This technique would remove the need for this, and would focus more plausibly on the relationship between less granular risk factors, for example between market risk and mortality. Again, judgement would be needed to decide the appropriate structure, including the number of layers and the calibration of the correlations within each matrix. Also, data can be used to derive the innermost correlation factors, but judgement will be needed for the outer layers, as the combined risk group no longer has a dataset behind it. Finally, where there are many layers, an equivalent single-layer structure cannot always be found. Both the single and the multi-layer approaches implicitly assume a linear relationship between the risk factors and the capital losses.

### 3.5.2 Capital-correlation not being the same as actual risk-correlation

An issue with covariance-correlation method is the correlation matrix used to aggregate capital usually does not reflect actual correlation between risk factors. Instead, capital-correlation reflects the risk profile of the insurer’s book of business. In the case of the Standard Formula, the correlation matrix is calibrated considering both actual risk-correlation and the exposure of an “average insurer” in the EU.

An example for this issue is double-sided capital-correlation, when a certain correlation factor effectively results in a reduced combined capital requirement compared to the combined capital requirement had the same factor been used but with the opposite sign. While this is a mechanism used to produce a prudent result, it is also a mechanism that would make it very difficult to explain why the correlation has changed sign (if it has) over successive periods of capital measurement. It also may be difficult to explain in the presence of ring-fenced funds, if opposite signs need to apply in determining the correlation factor inside a ring-fenced fund and the whole company that includes it. Also, where non-linearity exists, it may mean that the correlation between two risk factors, when a third risk factor is not stressed may have a different sign compared to when the third risk factor is stressed. This approach also does not reflect reality – the correlation between two risk factors will not generally depend on the exposure of an individual insurer to those risks.

For example, consider a company with an annuity portfolio where assets are of slightly longer duration than the liabilities, making it sensitive to a rise of interest rates. Say that this company is also at risk if equities fall. That would result in a “prudent” positive correlation between falling equities and rising interest rates. However, under a longevity stress, the portfolio could become susceptible to a fall of interest rates as the duration of assets becomes less than that of the liabilities, making the “prudent” positive correlation between falling equities and falling interest rates.

While interest rates is the most frequently quoted risk that attracts double-sided correlations, the issue could arise with other risks, particularly in the presence of ring-fenced funds where hedging strategies have been
employed (for example a fund may be susceptible to a rise in equities, while the company as a whole may be at risk from a fall).

### 3.6 Issues in calibrating the correlation matrices

Calibration and parameterisation of individual risk factors is not covered here. We are also not covering technical details for calibrating correlation factors, such as robustness of estimates, allowing for autocorrelation, identifying structural breaks in the data, etc. These matters are covered in a number of textbooks on Enterprise Risk Management (for example see [ref. SF, McNFE]); This is also further discussed in Section 4 and are beyond the scope of this paper.

We do want however to highlight some practical aspects of calibrating correlations for the covariance matrix method.

#### 3.6.1 Tail dependency

Tail dependency is an expression used to describe the extent to which the degree of association between risks is different in “benign” scenarios and in extreme scenarios. For example, while the correlation between mortality rates and credit default of reinsurers may be small under “normal” circumstances, it may be greater under scenarios where the mortality risk factors take extreme values; in such a scenario reinsurers may face calls from many insurers simultaneously and the risk of their default might be greater than normal. Given that capital requirements are defined in extreme scenarios, it is an important consideration in identifying capital requirements. Other examples may be returns of different types of assets or in different geographic locations becoming small/negative/very negative at the same time, due to market distress. Examples are the Black Monday in 1987 and the 2008-9 Global financial crisis.

This is discussed further in Section 4.7. Again, this is an area of increasing interest, also influenced by regulator views and significant judgement.

The covariance matrix approach can allow for tail dependency implicitly, by adjusting the correlation factor calculated directly by observing the data. However this implies that the model becomes less valid if used at other percentiles, which may restrict the purposes for which the model is used. Some companies may set different correlation assumptions in the body and the tail of the aggregate loss distribution which would require further validation.

#### 3.6.2 Coherent correlation factors

It is usually acceptable to use pairs of datasets of different lengths and frequency to determine correlations, which could mean that some correlation factors might be more reliable than others.

Basing correlations on a coherent dataset will result in a positive semi-definite (PSD) correlation matrix; if the dataset is not coherent (for example different parts of a dataset used to derive correlations between different risks) we may end-up with a non-PSD matrix which in turn may result in unintuitive results after aggregation, whereby the total diversified capital is more than the total undiversified capital (see [ref. SSS]).

Moreover, adjustments made to reflect tail dependency as well as exercising judgement in setting correlations may result in the correlation matrix not being PSD.

This topic is covered in more detail in section 4.

### 3.7 Limitations which remain & how to compensate for them

A key Solvency II requirement for firms that go through an IMAP is that the Internal Model can produce a probability distribution of losses, and for each point in the distribution we can determine the scenario (i.e. the combination of risk factors) that produced it. The full covariance approach requires significant effort to calculate losses at a specific percentile, after allowing for any intricacies of a particular insurer.
In theory, a distribution of losses can be extrapolated by fitting a distribution around the best-estimate balance sheet and the balance sheet at the 99.5\textsuperscript{th} percentile. A normal or lognormal distribution (or others) could hence be fitted. However, further runs of the heavy model may be required to validate this aggregate loss distribution together with a coherent justification of the level of the risk factors chosen to produce an aggregate loss at a different percentile. Identifying the combination of events (risk factors) to determine aggregate losses at a different percentile relies heavily on expert judgement. Such may be developed, after significant sensitivity analysis is performed over a number of iterations, and possibly by repeating the process at different times. It is therefore likely to be impractical for most insurers. Even then, significant effort may be needed to convince the regulator of the rigorousness and robustness applied to reach the Aggregate Loss Distribution.

Compared to the copula + proxy model approach, the choice of the distribution and its parameterisation are matters of judgement rather than an output of the model.

3.8 Key communication challenges

Allowance for the parts that are not naturally allowed for in the method are likely to cause most concern. Judgement is used to decide:

- How are risks grouped and why a particular multi-layered structure is appropriate
- How to allow for the difference between the capital aggregation correlation matrix and the actual risk-correlations, for example double-sided correlations
- How non-linear risks are allowed for
- How allowance is made for ring-fenced funds and any other transferability restrictions
- How allowance is made for tail dependency

3.9 Covariance Approach and Solvency II

Although not impossible, given the Solvency II requirements, it is difficult to build an internal model for Solvency II reporting solely based on the covariance matrix approach. As discussed in 3.7, the assumptions underlying the approach are unlikely to be realistic and simultaneously satisfy the Solvency II Statistical Quality Standards. It is possible however (and actually, to our knowledge, quite common) to have a model that uses a covariance matrix approach to model some risks as parts of its architecture but uses a simulation approach for the most critical risks to which the insurer is exposed. The fact that a sole covariance matrix approach can be challenging to defend for reporting purposes makes it also unlikely (yet possible) to be used for internal management, as ultimately the disclosed balance sheet is likely to be the one that drives the company’s decision making process. However the approach is still likely to be used as a validation tool to support results produced by other methods.

Proxy modelling techniques, including copula-generated scenarios are not new, however they have seen recently resurgence in their use in insurance, having been used previously in the banking industry as well as other non-financial applications. Advances in computing power also made it more feasible. The method is relatively simple to develop and a number of providers already exist that can market standard or tailored products to insurers. Of course, as we will see, simulation techniques still present limitations, however they address some of the key limitations of the covariance approach. Crucially they can produce a continuous aggregate loss distribution of losses and they can point to a combination of risks that result in a certain amount of loss, relatively easily (sometimes referred to as “what-if” scenarios).

Simulation techniques also make more transparent the assumptions made implicitly in a covariance matrix approach and provide the user with greater flexibility to depart from them. Of course, simulation techniques on this area are relatively new in the industry, so probably no standard market practice has been established in a number of areas.
4 Copulas

4.1 Introduction

A significant number of UK life assurers that plan on using an internal model to calculate their SCR under Solvency II intend to use a copula based simulation technique for their Risk Factor Model and a proxy model used for the Value Response Model (a “copula+proxy model” approach).

An illustration of this approach is provided in the diagram below.

There are several reasons why such an approach may be preferred to a correlation matrix approach:

- **More appropriate reflection of reality**
  The models allow a choice of model for risk factor changes and the ability to reflect non-linear balance sheet responses.

  Constraints on the ability of capital resources in one part of an undertaking to absorb losses elsewhere (“fungibility constraints”) and the effects of these constraints on diversification can be taken into account by modelling those constraints explicitly in each scenario. Alternatively these can be allowed for through the heavy model runs used to calibrate the proxy model.

  The techniques can also be used to model other exposures which are non-linear in nature (e.g. excess of loss reinsurance, management actions).

- **Transparency**
  It is straightforward to extract the simulations giving rise to a certain level of losses. This can provide insight into the kind of events that drive extreme losses and help in the validation of the model.

  It separates the model into natural and intuitive components. This facilitates identification of the most significant assumptions, allowing validators and other stakeholders to focus attention on these.

- **Reporting processes**
  The component structure can assist in acceleration of reporting timescales. For example, the risk factor model may be calibrated well in advance of a year end. The proxy model could be calibrated in advance of year end then rolled-forward and trued-up to reflect actual year end exposures.
If the proxy model is defined at an appropriate level of granularity, the simulated output from a single model run can be used to determine capital requirements for multiple business units (e.g. capital requirements for a group parent company and subsidiary insurance companies).

- **Use test**
  
  It produces a richer set of data. By producing a full distribution of outcomes, losses at different percentiles can be extracted from one set of simulation output without the need to re-run the model.
  
  The simulated losses can be produced separately for different portfolios of business and added up to obtain aggregate losses. This identification of losses at portfolio level allows capital requirements to be allocated those portfolios. This can help identify risk “hot spots”.
  
  The separation of the Risk Factor Model from the Value Response Model facilitates “what-if tests”. For example, the capital implications of the introduction of a change in hedging strategy or change in management actions could be assessed by amending the proxy model appropriately and re-running the simulations.
  
  It can support regular Solvency Monitoring. By rolling-forward the proxy model and re-basing the risk factor model to reflect changed conditions, the adjusted model can be re-run.

- **Meets regulatory requirements**
  
  It generates a full distribution of outcomes which can help meet the Internal Model requirements of Solvency II.

On the other hand, simulation based approaches have certain disadvantages:

- **Greater complexity**
  
  The modelling involved is significantly more complex and requires more assumptions and judgements to be made, adding to the cost of maintaining the model.

- **Potential reduction in transparency**
  
  Greater complexity can lead to the risk of models becoming more of a “black box” due to the larger number of assumptions and greater sophistication which may be more difficult for stakeholders to understand and critically evaluate.

Ultimately the choice of approach will depend on weighing up the relative advantages and disadvantages taking into account the nature of a firm’s own risk exposures and how it intends to use the model.

In the following sections, we focus on some of the main choices to be made when selecting and parameterising a copula. These choices can be significant. Often, the effects of diversification can be the largest single item presented in a breakdown of the SCR. The effects of diversification depend on the level of granularity chosen for such a breakdown. For example, whether the standalone capital requirements for each individual risk factor are presented separately or are combined to some intermediate level first (e.g. “market risk”, “insurance risk”). Industry surveys indicate that the effects of diversification typically amount to 40% to 60% of the sum of standalone capital requirements.

In practice, the majority of UK life insurers which are using a copula based approach have opted to use one of the elliptic family (e.g. Gaussian or Student’s t copula). One of the principal decisions to be made is whether to model tail dependence explicitly and how to account for it in the calibration of the copula. We provide some material which may be helpful in explaining the concept of tail dependence, both visually and in terms of the chances of the joint occurrence of extreme events. We also provide some background on statistical techniques for the fitting and testing of copulas and a worked example of their application in a life insurance context. We

---

7 For example, the 2015 Towers Watson Risk Calibration survey and the 2015 KPMG Technical Practices Survey. See [Ref. K].
believe that such techniques may be less familiar to actuaries working in life insurance. As one might expect, given the scarcity of data, the tests are less conclusive than in a one-dimensional setting. However, they may be helpful in evidencing compliance with statistical quality standards. Consideration of conditional probabilities also provides a useful framework for the calibration of a copula making allowance for tail dependence.

4.2 What is a copula and why is it useful?

At its most basic, a copula is a mathematical object that fully describes the dependency structure underlying a multivariate distribution.

It is a standard result (“Sklar’s Theorem”) that any multivariate distribution that describes changes in risk factors can be decomposed into the following:

- A set of one dimensional distributions that assign probabilities to changes in the individual risk factors (or “marginal distributions”);
- A copula that describes the dependency between those one dimensional distributions.

If the marginal distributions are continuous, the copula is uniquely determined.

Conversely, given a set of marginal distributions and a copula, you can construct a multivariate distribution which has those marginals and that copula.

It is the latter result which makes a copula useful for the purposes of capital modelling. By isolating the marginal distributions from the dependency structure, it facilitates a “bottom-up” approach to the construction of a joint distribution of risk factor changes:

- The first stage involves the selection of models for changes in each of the individual risk factors.
- The second stage then uses the copula to “glue” together samples from the changes in the individual risk factors in a way that reflects the assumed dependency structure. Perhaps the simplest way to think of a copula is an algorithm for matching up ranks (or, equivalently, quantiles) of the marginal distributions.

It is generally far easier to select and calibrate models for individual risk factors separately. This approach allows the modeller to focus on the data relevant to that risk factor and avoids the constraint of looking at shorter periods of time for which data for several risk factors is available. The fitting and validation of one dimensional models is also more straightforward. It is far easier to visualise data in low dimensions and so produce useful charts (e.g. QQ plots comparing actual outcomes with outcomes predicted by the model) to assist the demonstration of the appropriateness of the model. The selection and parameterisation of models for individual risk factors may also naturally fall to different sets of subject matter experts within the firm. For example, the modelling of insurance risks such as longevity and persistency generally involves a significant element of expert judgement and may be performed by personnel closely involved with the analysis of experience and selection of valuation and pricing assumptions. The modelling of economic risk factors such as equity returns and corporate bond spreads, which is generally more data driven, may fall more naturally to personnel with statistical and quantitative modelling skills.

There are also use test reasons for adopting a copula approach. The building block approach can provide greater transparency. Stakeholders may have views of what is reasonable for a “1 in 200 event” in an individual risk factor (say), providing for more robust validation and greater engagement. Risk appetites may be expressed in terms of capital requirements for individual risk factors. Splitting the model into more granular component may also assist in narrowing down and identifying the most significant assumptions and focussing validation effort on them.

Of course, isolating the dependency structure means that modellers, validators and their stakeholders also have to be satisfied that the assumptions made in respect of the copula model are reasonable!
4.3 Copula - definition

Mathematically, a d-dimensional copula is a cumulative distribution function on \([0, 1]^d\) with uniform marginal distributions; i.e. a function \(C : [0,1]^d \rightarrow [0,1]\) which satisfies the following conditions:

i. \(C(u_1, \ldots, u_d) = 0\) if \(u_i = 0\) for any \(i\)

ii. \(C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i\) for all \(i \in \{1, \ldots, d\}\); \(u_i \in [0,1]\)

iii. For each hyper-rectangle \(B = \prod_{i=1}^d [a_i, b_i] \subseteq [0,1]^d\) such that \(a_i \leq b_i\)

\[
\int_B dC(u) = \sum_{i=1}^d \sum_{j=1}^d (-1)^{i+j} C(u_{1:i}, \ldots, u_{d:b}) \geq 0
\]

where \(u_{1:i} = a_i\) and \(u_{2:j} = b_j\) for all \(i \in \{1, \ldots, d\}\)

Property (iii) ensures that \(Pr\{(u_1, \ldots, u_d) \in [a_1, b_1] \times \ldots \times [a_d, b_d]\} \geq 0\)

Sklar’s theorem (see also McNeil, Nelson)

Every multivariate distribution can be expressed in terms of a copula and its marginal distributions.

If \(F\) is a joint distribution function with continuous marginals \(F_{1\ldots d}\), then there is a unique copula \(C\) such that:

\[
F(x_1, \ldots, x_d) = C\left(F_1(x_1), \ldots, F_d(x_d)\right) \text{ for each } (x_1, \ldots, x_d) \in \mathbb{R}^d
\]

Conversely, given a copula \(C\) and continuous univariate distribution functions \(F_1, \ldots, F_d\), the multivariate distribution \(F\) defined by (##) has marginals \(F_1, \ldots, F_d\)

For a continuous multivariate distribution \(F\), its copula is defined by the mapping

\[(u_1, \ldots, u_d) \in [0,1]^d \rightarrow C(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)) \in [0,1]\]

4.4 Using copulas in simulations

4.4.1 Inverse distribution function method

An algorithm for simulating values \((u_1, \ldots, u_d)\) from the copula can be thought of as a recipe for generating d-tuples\(^8\) of quantiles (or ranks) in a way which respects the dependency structure. Provided we have such a recipe, it defines an algorithm for generating simulations of the \((X_1, \ldots, X_d)\):

\[(u_1, \ldots, u_d) \in [0,1]^d \rightarrow (X_1, \ldots, X_d) = \left(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right)\]

Each simulated d-tuple of quantiles simulated from the copula “picks out” corresponding quantiles from the marginal distributions and “glues them together” into a corresponding d-tuple of changes in risk factors.

This approach forms the basis of the usual “inverse distribution function method” of implementing a copula illustrated for the case \(d=2\) in the diagram below.

---

\(^8\) A d-tuple is a sequence / ordered list of ‘d’ elements, where ‘d’ is a non-negative integer.
In this case, we have an algorithm which produces simulated pairs of values \((U_1, U_2)\) where each of \(U_1\) and \(U_2\) take values in the interval \([0, 1]\). The simulated values are illustrated as a scatter plot in the top right hand corner of the diagram. This is a typical way of illustrating a copula.

We have also defined models for the two individual risk factors \(X_1\) and \(X_2\). These models are defined by (cumulative) distribution functions which assign probabilities to \(X_1\) and \(X_2\). For example, the probability that \(X_1\) takes a value less than \(x_1\) is given by \(F_1(x_1)\).

Our simulation recipe maps \(U_1\) to \(X_1 = F_1^{-1}(x_1)\) and \(U_2\) to \(X_2 = F_2^{-1}(x_2)\).

It then glues them together to form the pair \((X_1, X_2)\).

The result is a large number of simulated values of \((X_1, X_2)\) which have the dependency defined by the copula.

The approach generalises in an obvious way for higher dimensions than two (i.e. \(d>2\)). However, it becomes much harder to visualise scatter plots of copulas in higher dimensions.
4.4.2 The Iman-Conover method

The inverse distribution function method of implementing a copula starts with simulated values from the copula and uses the inverse of the distribution function to identify values of the marginal random variables.

The Iman-Conover method provides an alternative approach for the implementation of a copula which can be particularly useful where simulated values of the marginal random variables have already been generated. As was noted above, a copula defines a relationship between the quantiles or ranks of the marginal random variables. Given a set of N simulations of d-tuples of random variables, the Iman-Conover method first forms these into an Nxd matrix, each of whose rows is one of simulated d-tuples. The method then proceeds by forming another Nxd matrix from N simulations of the chosen d-dimensional copula. Each of the N rows of this “copula matrix” corresponds to one of the simulations from the d-dimensional copula and all entries take values in the interval [0,1]. This copula matrix is then transformed into an Nxd matrix of ranks by mapping each observation in each column to its rank in that column (i.e. an integer value between 1 and N). The entries in each of the d columns of the original Nxd matrix of simulated values from the marginal distributions are then shuffled so that their ranks match those of the transformed copula matrix.

The Iman-Conover method therefore re-orders the simulated values of the marginal distribution so that their ranks match the pattern produced by the copula. It therefore provides a more intuitive implementation of a copula which may be easier to explain to stakeholders.

An illustration of the Iman-Conover method in the case N=5, d=2 is provided below.

Copula orders ranks of the marginal distributions to produce the multivariate distribution using the Iman-Conover method.

<table>
<thead>
<tr>
<th>#</th>
<th>1. Marginal distributions (independent)</th>
<th>2. Copula</th>
<th>3. Multivariate distribution (reordered marginals)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Values</td>
<td>Ranks</td>
<td>Values</td>
</tr>
<tr>
<td>1</td>
<td>Y1</td>
<td>Y2</td>
<td>Y1</td>
</tr>
<tr>
<td>2</td>
<td>-1.70</td>
<td>0.06</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>-0.47</td>
<td>-1.18</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>9.13</td>
<td>-2.54</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>-3.42</td>
<td>1.27</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0.15</td>
<td>0.69</td>
<td>4</td>
</tr>
</tbody>
</table>

Figure 4-3 Copula Orders Ranks Producing Multivariate Distribution

Readers are referred to the original paper of Iman and Conover for more details. A particularly accessible account is provided by Mildenhall. Mildenhall describes how the method can be implemented using techniques of linear algebra for the Gaussian copula and extended to the Student’s t copula. However, the method can be implemented for any copula – all that is required is the Nxd matrix whose N rows are simulated values from the d-dimensional copula.

---

9 Iman-Conover, A distribution-free approach to inducing rank correlation among input variables, 1982 (See [ref. IC])

10 Correlation and Aggregate Loss Distributions with an Emphasis on the Iman-Conover Method, Stephen J. Mildenhall, 2005. (See [ref. MiS])
4.5 Choosing a simulation method

In practice, the choice of using the inverse distribution function or Iman-Conover method will be made for the end-user if buying a proprietary aggregation tool. As they are equivalent approaches to implementing a copula, one may be neutral. For example, even if one has defined a marginal distribution in terms of simulated values (e.g. from a time-series forecast), it is possible to produce an empirical distribution function from the simulated values by fitting an appropriate function (e.g. a piecewise linear function).

While both methods of implanting a copula are equivalent, the Iman-Conover method may be easier to code and avoids the intermediate step of estimating the CDF, thus making sensitivity testing easier when applying different correlation structures to these same marginal distributions.

4.5.1 Selecting and parameterising a copula

When selecting a copula model and its parameterisation, life insurance companies are typically faced with a number of considerations which may influence their choice of model:

4.5.2 Number of risk factors

Life insurers are typically faced with a number of risks:

- Market risks (e.g. equity returns, interest rates, property returns, foreign exchange risk, bond spreads, inflation, equity implied volatility …)
- Counterparty default risks (corporate bonds, cash and derivative counterparties, reinsurance counterparties…)
- Insurance risks (annuitant longevity, life assurance mortality, persistency, expense …)
- Operational
- Liquidity

They may be exposed to these risks across a number of different business units or geographical territories.

Depending on the nature and complexity of the risks, the number of risk factors included within a model can become quite large, particularly for multi-national groups which write a diverse range of business. This has implications for the nature of the copula model.

4.5.3 Data

The availability of relevant data may influence the choice of copula. The greater the volume of data, the more one may be able to rely on statistical methods to guide the choice of a copula and its calibration.

In the context of life assurance, relevant data is typically very limited and any kind of meaningful statistical analysis is usually only possible for market risks, if at all. The length of time periods over which separate data series coincide is even more limited. This leads to the selection of assumptions based on non-coincident subsets of data in order to maximise the data used. Even then, there remains significant uncertainty. As those familiar with setting correlation assumptions are aware, different periods of time can suggest different assumptions. The selection of models and assumptions therefore requires a significant overlay of expert judgement. This piecemeal process of assumption selection guided by expert judgement can lead to internal inconsistencies within the initial parameterisation which requires subsequent adjustment (e.g. an initially chosen correlation matrix may not be internally consistent, as discussed in section 4.10).

4.5.4 Tail dependence

Whether or not to allow for tail dependence explicitly in the choice of a copula model depends on modelling preferences. If one believes strongly that tail dependence is a feature that should be explicitly captured, then this may rule out certain models such as the Gaussian. Alternatively, one may choose a less complex model such as a Gaussian but adjust the parameterisation to make an allowance for tail dependence. We discuss tail dependence in section 4.7.
4.5.5 Heterogeneity

Heterogeneity in the types of risk factor modelled makes it less straightforward to form prior views on aspects of modelling such as tail dependence. Whereas one may expect tail dependence between returns on equities of different companies operating in a given sector due to the presence of common factors which influence the values of all shares in that sector, it is less clear whether there is tail dependence between risks such as changes in the level of interest rates and mis-estimation of persistency rates.

4.5.6 Through the cycle vs point in time

The issues to be considered here are similar to those for the individual risk factor distributions and the choice will usually be determined by a company’s more general philosophy regarding capital modelling.

- Avoidance of pro-and counter-cyclical effects.
  One may wish to avoid triggering actions during times of stress or releasing capital during benign periods.
- Practicality
  Correlation matrices underlying copulas form a large set of assumptions. It may not be a practical proposition to review all the assumptions on an annual basis. Judgements may also not change frequently, particularly where correlations involve non-market risks. The extent to which one should rely on very recent data when selecting correlations between market risks is also a consideration. One may prefer a less frequent regular review assumptions with ad-hoc reviews triggered if there is judged to have been a significant market dislocation.

4.5.7 Use test and transparency

The more complex a model, the greater the risk of it becoming a “black box”. The more complex and less intuitive the model, the more it is difficult to identify and understand the assumptions that most significantly affect the model outputs. It is important for users to understand the levers they have to pull and how they affect the figures produced. Models should be no more complicated than they need to be. It is also important that users understand the limitations of the model. In some cases, it may be appropriate to adopt a simpler, more transparent model whilst making allowances for its limitations.

4.5.8 Practicalities – standard packages

A proprietary aggregation tool may come with a number of copulas as standard. In the event that users wish to adopt a non-standard model, this may require the development of an in-house tool unless the proprietary package is sufficiently flexible to allow input of simulated copula data.

4.6 Types of copula

Copulas come in three main categories:

- Implicit copulas
- Explicit copulas
- Empirical copulas

Implicit and explicit copulas are parametric models and are covered extensively by academic literature. If one has views on the properties that a copula should have, one can look through the families of models in the literature to find suitable candidates.

In this section, we provide overviews of these classes of copulas and their properties.

4.6.1 Implicit copulas

Implicit copulas are copulas which underlie standard multivariate distributions (e.g. the Gaussian or Student’s t copula). Their distribution functions do not typically have simple closed form expressions. However, some of the models are relatively straightforward to simulate from with the simulation algorithm exploiting some of the linear algebraic properties of the multivariate distribution from which the copula is derived.
For example, a d-dimensional Gaussian copula is defined by a positive semi-definite matrix $\Sigma$. An algorithm for simulating from the copula is as follows:

1. Find a matrix $A$ such that $\Sigma = A^T A$ (e.g. a lower triangular matrix found by Cholesky decomposition or an orthogonal matrix of eigenvectors using spectral decomposition)
2. Generate N pseudo-random simulations of d-tuples of independent Normal random variables. For example, generate N d-tuples of random $U(0,1)$ variables $(U_1^{(k)}, \ldots, U_d^{(k)})$ for $k = 1, \ldots, N$ and transform into d-tuples of random $N(0,1)$ variables using the transformation $(U_1, \ldots, U_d) \rightarrow (Z_1, \ldots, Z_d) = (\Phi^{-1}(U_1), \ldots, \Phi^{-1}(U_d))$.
3. For each simulation, form a d-tuple of dependent Normal random variables $X^{(k)} = AZ^{(k)}$.
4. Convert the dependent Normal random variables back to “copula space” $[0, 1]^d$ using the transform $(X_1, \ldots, X_d) \rightarrow (V_1, \ldots, V_d) = (\Phi(X_1), \ldots, \Phi(X_d))$

The algorithm for a d-dimensional Student’s t copula defined by a positive semi-definite matrix $\Sigma$ with degrees of freedom parameter $\nu$ is similar:

1. Generate N pseudo-random simulations of d-tuples of dependent Normal random variables $(X_1, \ldots, X_d)$ as above.
2. Generate independently N simulations of $U(0,1)$ variables.
3. Set $W = G_{\nu}^{-1}(U)$ where $G_{\nu}$ is the distribution function of $\chi^2_{\nu}$.
4. Form N d-tuples $(Z_1, \ldots, Z_d)$ by rescaling the vectors generated in step 1: $(Z_1, \ldots, Z_d) = \left( \frac{\nu}{\nu W_1^{1/2}} X_1, \ldots, \frac{\nu}{\nu W_d^{1/2}} X_d \right)$ where $t_{\nu}$ is the distribution function of a 1-dimensional Student’s t random variable with $\nu$ degrees of freedom.

The Gaussian copula is defined uniquely by the correlation matrix $\Sigma$. It therefore has $d(d - 1)/2$ parameters – one for every pair of marginal distributions. This allows the association between each pair of variables to be controlled separately (subject to the correlation matrix remaining internally consistent or PSD).

The Gaussian copula is an example of a so-called elliptic copula. The probability density function of the distribution from which it is derived has elliptic level sets – see Figure 4.4.

It is a well-known fact that the coefficients of tail dependence of a Gaussian copula are zero. The lack of explicit tail dependence is a limitation of the Gaussian copula which, in some applications may result in adjustments to the correlation parameters. Tail dependence is discussed further in section 4.7

The Student’s t copula is another example of an elliptic copula. It is defined by $1 + d(d - 1)/2$ parameters.

The effect of the scaling (or “mixing”) variable used in the simulation algorithm is to introduce a common shock across all the risk factors. The presence of this common underlying risk driver leads to a tendency for all risk factors to take large or small values at the same time and the presence of tail dependence. The coefficients of tail dependence of a Student’s t copula are non-zero and are given by $2 t_{\nu+1} \left[ -\sqrt{\nu + 1}(1 - \rho)/(1 + \rho) \right]$ where $\rho$ is the parameter of the correlation matrix corresponding to the two variables. The possible coefficients of tail dependence form a space of $1 + d(d - 1)/2$ dimensions. However there are $d(d - 1)/2$ pairs of variables. It would require $d(d - 1)/2$ parameters to separately control both correlation and tail dependence for each possible pair of random variables. The limited number of free parameters therefore acts as a constraint on parameterising the copula, though, depending on risk exposures, it may not be significant.

The simulation algorithm for the Student’s t-copula generalises in a natural way that provides a greater number of free parameters. The steps from step (3) onward are replaced by:

3. Set $W_j = G_{\nu_j}^{-1}(U)$ where $G_{\nu_j}$ is the distribution function of $\chi^2_{\nu_j}$, for $j = 1, \ldots, d$.
4. Form N d-tuples $(Z_1, \ldots, Z_d)$ by rescaling the vectors generated in step 1: $(Z_1, \ldots, Z_d) = \left( \frac{\nu_1}{\nu_1 W_1^{1/2}} X_1, \ldots, \frac{\nu_d}{\nu_d W_d^{1/2}} X_d \right)$
(5) Set \((V_1, \ldots, V_d) = (t_\nu(z_1), \ldots, t_\nu(z_d))\) where \(t_\nu\) is the distribution function of a 1-dimensional Student’s t random variable with \(\nu\) degrees of freedom.

The resulting copula is the so-called “individuated” or “generalised” Student’s t copula or IT copula. Like both the Gaussian and Student’s t copula, it is straightforward to simulate from.

The presence of the common shock variable results in tail dependence. However, there are now \(d\) additional parameters which can be used to control tail dependence compared to just 1 for the Student’s t copula. However, the \(d\) additional parameters are not sufficient to fully control the coefficients of tail dependence for the \(d(d-1)/2\) pairs of risk factors. Unfortunately, the formula for the coefficient of tail dependence does not have quite such a compact expression as that for the Student’s t copula – (see [ref. SL1] for more details).

The IT copula is also no longer an elliptic copula (except in the special case where all the degrees of freedom parameters are equal). However, it is “radially symmetric” – i.e. symmetric along rays drawn through the centre of the hypercube \([0, 1]^d\). This means that the IT copula is capable of exhibiting skewed behaviour.

A Grouped T-copula is a special case of the IT copula in which the degrees of freedom parameters are equal for certain blocks i.e. \(\nu_1 = \nu_2 = \cdots \nu_{i_1}, \nu_{i_1+1} = \nu_{i_1+2} = \cdots \nu_{i_2}, \ldots, \nu_{i_k+1} = \nu_{i_k+2} = \cdots \nu_d\).

The diagrams below show
- Scatter plots of a bivariate distribution formed using the Gaussian, Student’s t and IT copulas with standard Normal marginal distributions.
- Contour and 3D plots of the probability density function
- The same plots but for the copula (i.e. on a quantile scale)

The elliptic symmetry of the Gaussian and Student’s t copula can be seen from the contour plots. The contour plot of the IT copula shows using different degrees of freedom assumptions can produce a skew in the distribution (in this case, towards the NNE and SSW octants).

The presence of tail dependence can also be seen in the Student’s t copula from the clustering of points in the NE and SW corners. For the IT copula, the clustering is more evident in the edges of the NNE and SSW octants. The presence of tail dependence is also indicated by the curvature of the Student’s t copula becoming more “pointed”.

The clustering of points is more evident in the scatter plots of the copula. The density function of the Student’s t copula has a more peaked “tail”. The skewness of the IT copula is illustrated by the raising of the sheet in the NNE and SSW octants.
Figure 4-4 Scatter / Contour Plot of probability density function for bivariate meta-distribution with $N(0,1)$ marginals

Figure 4-5 Perspective Plot of probability density function for bivariate meta-distribution with $N(0,1)$ marginals
4.6.2 Explicit copulas

Explicit copulas have closed form expressions for their distribution function and include the widely studied class of Archimedean copulas. The Archimedean copulas contain several well-known families of copulas including the Gumbel, Clayton and Frank copulas. The academic literature contains lists of such copulas (see, for example, Joe and Nelson).
The distribution functions of these copulas typically have a relatively small number of parameters (up to 3) to control the “shape” of the copula. Depending on the value of the parameters, the copulas can exhibit varying degrees of tail dependence. The relatively small number of parameters means such copulas are likely to be less useful in describing the relationships within a high dimensional set of risk factors. The low number of free parameters also results in a high degree of symmetry. Perhaps for these reasons, they have found relatively little application in life insurance to date. However, it possible that they could find application in future through the use of “vine copulas” – see section 4.6.4.

Surprisingly given the closed form expression for the distribution function, Archimedean copulas can be difficult to simulate from. This contrasts with the elliptic family of implicit copulas where the simulation algorithm is straightforward but the formula for the copula distribution function is not.

The charts below show scatter and contour plots for two standard bivariate Archimedean copulas (the Clayton and Gumbel copula) and the corresponding meta-distributions with standard Normal marginals. The standard versions of both copulas are defined by a single parameter (“theta”). We have chosen theta to give a Spearman’s rank correlation of around 70%.

The Clayton copula has a positive coefficient of lower tail dependence. Its coefficient of upper tail dependence is zero.

The Gumbel copula has a positive coefficient of upper tail dependence. Its coefficient of lower tail dependence is nil.

These features can be observed in the clustering of points in the tails of the scatter plots and the pointedness of the contours of the meta-distributions.

Figure 4-8 Meta-Clayton versus Meta-Gumbel Marginals (theta = 2.1)
Figure 4-9 Meta-Clayton (theta = 2.2) versus Meta-Gumbel Marginals (theta = 2.0)

Figure 4-10 Clayton (theta = 2.1) versus Gumbel (theta = 2.1)
Empirical copulas are derived from sample data. As we have seen above, parametric copulas generally exhibit symmetry in some form. Being based on data, an empirical copula can more closely reflect any lack of symmetry or any other feature of the data set used.

If there are M data points in total, each consisting of a d-tuple of risk factor observations, a random sample can be made from an empirical copula as follows:

(i) Assign each of the data points a unique label from 1 to M;
(ii) generate a pseudo-random number between 1 and M and identify the corresponding data point;
(iii) taking each coordinate separately, calculated the rank of the observed value and divide by (M+1)
(iv) This produces a d-tuple, each of whose coordinates takes a value in [0,1].

This approach will produce only a discrete number of distinct observations. The approach described above can be refined by segmenting the unit interval resulting in a division of the hypercube [0,1]^d into a “grid” of “mini-hyper-cubes”. The procedure described above is followed and the “mini-hyper-cube” into which the simulation generated at step (iv) falls is then identified. The algorithm then generates another d pseudo-random U[0,1] random variables and uses these to linearly interpolate between the edges of the “mini-hyper-cube” to produce the final simulated value.

However, there are a number of points which are relevant in a life insurance context:

- An empirical copula is only as good as the data which is available. Simulating from an empirical copula can extract no more information than that which is found in the sample data.
- Even where data is available, the sample points used for the empirical copula have to be populated with coordinates in each risk factor. The data for each risk factor therefore needs to be sampled from the same points in time. This limits the use of data to the longest period for which data is available for all risk factors to be modelled.
- An empirical copula is essentially a step function. The scantiness of data inevitably means that the empirical copula does not provide a very rich structure.
- Use of an empirical copula is simply a way of extrapolating the relationships observed in the data. There may be reasons why relationships observed in the past may not be considered entirely appropriate for the modelling of risk in the future. It is more straightforward to adjust for such views if using a parametric copula.

Figure 4-11 Clayton (theta = 2.1) versus Gumbel (theta = 2.1) Marginals

4.6.3 Empirical Copulas
There are means of overcoming these limitations by using synthetic data produced by a model rather than actual historical observations. For example, one could use a real world Economic Scenario Generator to generate a large number of simulations of changes in market risk factors which could then be used to derive an empirical copula. An extended ESG which has been augmented to include non-market risk factors could be used if one wished to use an empirical copula that covered all risk factors. However the additional step to calibrate and simulate from the real world ESG may reduce the tractability and transparency of the copula.

Results of industry benchmarking surveys suggest that empirical copulas (including those based on synthetic data) are currently not commonly used within the UK life insurance industry.11

### 4.6.4 Vine copulas

As we have seen above, copulas are far easier to visualise in low dimensions. They are also more straightforward in low dimensions where there are more likely to be longer periods of contemporaneous data. Vine copulas (see, for example, [ref. A, BS, and J]) provide a means of joining copulas, albeit at the cost of additional complexity. The Working Party has not considered vine copulas in any detail and their use currently within the UK life insurance industry does not appear to be common. However, we include a reference here as a possible area for future investigation.

### 4.6.5 Current practice

According to the 2015 Risk Calibration Survey conducted by Towers Watson, 10 of the 28 respondents use a copula based approach for their capital models. The remaining companies use a correlation matrix or, in one case, a mixed approach. Of the 10 companies that use copulas, 7 use a Gaussian copula and 3 use a Student’s t copula.

The authors are not aware of any UK life insurer that has adopted an approach based on explicit, empirical or vine copulas. The reference to a “mixed approach” may infer a combination of a copula and covariance matrix approach or, possibly, a more complex approach.

### 4.6.6 Summary table

The table below provides a brief summary of the differences between the three types of copula.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Explicit</th>
<th>Implicit</th>
<th>Empirical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examples</td>
<td>Clayton</td>
<td>Gaussian</td>
<td>Actual data</td>
</tr>
<tr>
<td></td>
<td>Gumbel</td>
<td>Student’s t</td>
<td>Synthetic data</td>
</tr>
<tr>
<td></td>
<td>Frank</td>
<td>Grouped T</td>
<td>(e.g. from a real world ESG)</td>
</tr>
<tr>
<td></td>
<td>Joe</td>
<td>Individuated T</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ali-Mikhail-Haq</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is the Copula parametric?</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Can it be extended to many risk drivers?</td>
<td>Limited</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Number of parameters</td>
<td>Description</td>
<td>Usually one or two, some have up to 4 parameters</td>
<td>One per pair of risk drivers) + further parameter</td>
</tr>
</tbody>
</table>

4.7 Tail Dependence

4.7.1 Introduction

Tail dependence is a measure of the likelihood of a coincidence in extreme changes in two or more risk factors i.e. of the probability of “bad things happening together”.

One of the most important decisions to be made when selecting a copula and its parameters is whether or not to choose a model that exhibits tail dependence and how to allow for tail dependence in its calibration.

4.7.2 Definition

The mathematical definition of tail dependence is given in terms of conditional probabilities; i.e. given that an extreme change in one or more risk factors has occurred, what is the probability that an extreme change in another risk factor will occur?

The usual measures of tail dependence – upper and lower coefficient of tail dependence - are asymptotic measures. They are limiting values of the coefficient of finite tail dependence (see [ref. SF]) and it is perhaps easier to explain the concept of tail dependence in terms of this latter measure as it permits the use of graphical tools such as scatter plots. We will also see later in section 4.9 that charts of the coefficient of finite tail dependence can be a useful tool in informing the choice of copula parameters.

The coefficients of upper or lower finite tail dependence are functions, \( \lambda_U(q) \) and \( \lambda_L(q) \), defined as follows:

\[
\lambda_U(q) = \Pr((F_X(X) > q | F_Y(Y) > q) \\
\lambda_L(q) = \Pr((F_X(X) < (1 - q) | F_Y(Y) < (1 - q))
\]

i.e. the coefficient of finite upper tail dependence function \( \lambda_U(q) \) is the probability that a value of X exceeds the \( q^{th} \) percentile of X given that a value of Y has been observed that exceeds the \( q^{th} \) percentile of Y. \( \lambda_U(q) \) is a measure of the probability that X takes an extreme high value given that Y takes an extreme high value.

<table>
<thead>
<tr>
<th>Ease of simulating values</th>
<th>Technique</th>
<th>Exhibits tail dependence?</th>
<th>N – Gaussian</th>
<th>N – but produces conditional probabilities consistent with data used</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Upper Tail</td>
<td>Gumbel, Generalised Clayton, Joe</td>
<td>As above</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lower Tail</td>
<td>Clayton, Generalised Clayton</td>
<td>As above</td>
</tr>
</tbody>
</table>

| (for \( n \) risk drivers) | \( \approx 2 \) | \( \approx n^2 \) | n/a |
Similarly, $\lambda_L(q)$ is the probability that a value of $X$ is less than the $(1-q)^{th}$ percentile of $X$ given that a value of $Y$ has been observed that is less than the $(1-q)^{th}$ percentile of $Y$. $\lambda_L(q)$ is a measure of the probability that $X$ takes an extreme low value given that $Y$ takes an extreme low value.

Note that both $\lambda_U(q)$ and $\lambda_L(q)$ are probabilities (not correlations) and so take values in the interval $[0,1]$. For radially symmetric (2 point symmetry) copulas (such as the Gaussian, T and IT), the coefficients of finite upper and lower tail dependence are equal.

Visually this can be seen as:

<table>
<thead>
<tr>
<th>Upper Tail (Y vs X)</th>
<th>Lower Tail (Y vs X)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Upper Tail Graph" /></td>
<td><img src="image" alt="Lower Tail Graph" /></td>
</tr>
</tbody>
</table>

$\lambda_U(q)$ is the proportion of observations falling in the dark blue shaded box (upper right corner) divided by the proportion of observations falling in the lighter blue shaded rectangle.

$\lambda_L(q)$ is the proportion of observations falling in the dark blue shaded box (lower left corner) divided by the proportion of observations falling in the lighter blue shaded rectangle.

Note that the dark blue shaded box is part of the lighter blue shaded rectangle in both cases.

Figure 4-12 Upper and Lower Finite Tails

The coefficients of upper and lower tail dependence $\lambda_U$ and $\lambda_L$ are limiting values of coefficients of finite tail dependence given by:

$$\lambda_U = \lim_{q \to 1^-} \lambda_U(q)$$

$$\lambda_L = \lim_{q \to 1^-} \lambda_L(q)$$

We have used a slightly different definition of the coefficient of lower finite tail dependence than the usual one in order to be able to show both the upper and lower coefficients on the same axis.

Figure 4-13 illustrates the coefficient of upper tail dependence. As the events become more extreme (i.e. $q$ increases), the square in upper right hand corner shrinks, i.e., as we move from the light blue square to beige to dark grey. The coefficient of tail dependence is the limiting value of the ratio of the proportion of events that occur in the shrinking square to the proportion of events that occur in the shrinking rectangle.
The clustering of events in the tails is an indicator of tail dependence. Figure 4-14 shows a scatter plot from a bivariate Gaussian copula and a Student’s t copula. Both copulas have a correlation coefficient of 0.5. The Student’s t copula has 4 degrees of freedom.
It is a standard result that the coefficients of tail dependence of a Gaussian copula are zero while those of a Student’s t copula are non-zero (ref. McNeil). It is apparent from Figure 4-14 that the Student’s t copula shows a greater clustering of points in the squares in the upper right hand corner and lower left hand corner when compared to the Gaussian copula.

4.7.3 Tail dependence and correlations

The charts below illustrate the behaviour of the coefficient of finite tail dependence for the bivariate Gaussian copula and a Student’s t copula with 5 and 10 degrees of freedom. This is produced for three different correlation assumptions. (The right hand plot in Figure 4-14 is restricted to events more extreme than the 90th percentile.)

The following is apparent from the charts:

- The correlation parameter is the principal factor that determines conditional probabilities
- There is little difference in conditional probabilities in the body of the distribution
- However, differences do become apparent in the extreme tails
- It is clear that the coefficients of finite tail dependence tend to zero for the Gaussian copula but to non-zero values for the Student’s t copula.

Figure 4-15 Coefficients of Finite Tail Dependence, correlations of 25%, 50% and 75% (Expanded view 90th percentile on the right)

A coefficient of tail dependence which is equal to zero does not mean that the corresponding risk factors are “independent in the tail” (see, for example, MS).

Nor does it imply that extreme changes in one risk are less likely to be accompanied by extreme changes in another. In the case of the Gaussian copula, a positive correlation assumption between increases in X and increases in Y means that, on average, large increases in Y will tend to occur if a large increase in X as occurred. Recall that the conditional distribution is given by:

\[ Y|X \sim N(\mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (X - \mu_X), (1 - \rho^2)\sigma_Y^2) \]
### 4.7.4 Significance of tail dependence

When explaining the significance of tail dependence to stakeholders and why it is something that they should take into account, it may be useful to provide some simple quantitative indicators of what different copula models mean for the likelihood of “bad things happening at the same time.”

The tables below show:

- A comparison of joint exceedance probabilities at differing percentiles produced by a Student’s t copula with those produced by a Gaussian copula for various correlation and degrees of freedom parameters.

The table shows the probabilities for the Gaussian copula and the factors by which those probabilities must be multiplied to obtain the corresponding probabilities for the Student’s t copula. For example, assuming a correlation parameter of 50%, the probability that both risk factors exceed their “1 in 100 year” values at the same time under a Student’s t-copula model with 5 degrees of freedom is twice that under a Gaussian copula model. An event with a probability of 1 in 770 years under the Gaussian copula now has a probability of 1 in 385 years under the Student’s t copula. See the yellow highlighted cells below.

- A comparison of joint exceedance probabilities for d-tuples of risk factors to illustrate how tail dependence influences behaviour in higher dimensions. The tables show the joint exceedance probabilities for the Gaussian copula and corresponding multiples for the Student’s t copula with various degrees of freedom. We show values for 2, 5, 10 and 25 dimensions and at the 75th and 90th percentiles (to illustrate dependence on event severity).

For example, taking 90th percentile (or “1 in 10 year” events) in each risk factor and assuming a correlation parameter of 25%, a simultaneous event in 10 risk factors each of which is at least as severe as a 1 in 10 year event, is 5.7 times more likely if they follow a Student’s t module with 5 degrees of freedom compared to a Gaussian. The equivalent multiplier for 2 dimensions is 1.27. See the yellow highlighted cells below in the table using 25% correlation.

It should be noted in the below tables that the figures for the Student’s t distribution with degrees of freedom 2, 5, 10 and 30 are provided as multiples relative to the figures for the Gaussian distribution. In contrast, the figures for the Gaussian distribution can be read from the table directly. This allows easier comparison between the Student’s t and Gaussian distribution.

---

Table 4-2 Joint exceedance prob. (Gaussian and bivariate Student’s t copula shown as multiples of the Gaussian)

<table>
<thead>
<tr>
<th>Percentile</th>
<th>75th</th>
<th></th>
<th></th>
<th></th>
<th>90th</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correlation/DOF</td>
<td>2</td>
<td>5</td>
<td>10</td>
<td>30</td>
<td>Gaussian</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>0%</td>
<td>1.18 1.07 1.04 1.01</td>
<td>6.25%</td>
<td>2.18 1.50 1.25 1.08</td>
<td>1.00%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25%</td>
<td>1.11 1.04 1.02 1.01</td>
<td>8.93%</td>
<td>1.63 1.27 1.14 1.05</td>
<td>1.93%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td>1.06 1.03 1.01 1.00</td>
<td>12.03%</td>
<td>1.34 1.15 1.08 1.03</td>
<td>3.24%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75%</td>
<td>1.03 1.01 1.01 1.00</td>
<td>15.93%</td>
<td>1.17 1.07 1.04 1.01</td>
<td>5.12%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95%</td>
<td>1.01 1.00 1.00 1.00</td>
<td>20.98%</td>
<td>1.05 1.02 1.01 1.00</td>
<td>7.79%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>95th</th>
<th></th>
<th></th>
<th></th>
<th>99th</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correlation/DOF</td>
<td>2</td>
<td>5</td>
<td>10</td>
<td>30</td>
<td>Gaussian</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>0%</td>
<td>3.97 2.24 1.61 1.20</td>
<td>0.25%</td>
<td>18.46 7.45 3.72 1.74</td>
<td>0.01%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25%</td>
<td>2.38 1.61 1.31 1.10</td>
<td>0.61%</td>
<td>6.30 3.33 2.12 1.35</td>
<td>0.04%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td>1.69 1.32 1.16 1.06</td>
<td>1.22%</td>
<td>3.05 2.00 1.52 1.17</td>
<td>0.13%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75%</td>
<td>1.31 1.15 1.08 1.03</td>
<td>2.20%</td>
<td>1.78 1.41 1.22 1.08</td>
<td>0.32%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95%</td>
<td>1.09 1.04 1.02 1.01</td>
<td>3.71%</td>
<td>1.20 1.11 1.06 1.02</td>
<td>0.67%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4-3 Joint exceedance probabilities at 75th percentile

<table>
<thead>
<tr>
<th>Correlation</th>
<th>0%</th>
<th>25%</th>
</tr>
</thead>
<tbody>
<tr>
<td>dim/DOF</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>1.18</td>
<td>1.07</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>1.83</td>
</tr>
<tr>
<td>10</td>
<td>35</td>
<td>9</td>
</tr>
<tr>
<td>25</td>
<td>2.21E+05</td>
<td>8.56E+03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Correlation</th>
<th>50%</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>dim/DOF</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>1.06</td>
<td>1.03</td>
</tr>
<tr>
<td>5</td>
<td>1.22</td>
<td>1.09</td>
</tr>
<tr>
<td>10</td>
<td>1.37</td>
<td>1.15</td>
</tr>
<tr>
<td>25</td>
<td>1.64</td>
<td>1.26</td>
</tr>
</tbody>
</table>

Table 4-4 Joint exceedance probabilities at 90th percentile

<table>
<thead>
<tr>
<th>Correlation</th>
<th>0%</th>
<th>25%</th>
</tr>
</thead>
<tbody>
<tr>
<td>dim/DOF</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>2.18</td>
<td>1.50</td>
</tr>
<tr>
<td>5</td>
<td>70</td>
<td>16.50</td>
</tr>
<tr>
<td>10</td>
<td>66,030</td>
<td>4.239</td>
</tr>
<tr>
<td>25</td>
<td>3.50E+14</td>
<td>1.69E+12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Correlation</th>
<th>50%</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>dim/DOF</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>1.34</td>
<td>1.15</td>
</tr>
<tr>
<td>5</td>
<td>2.26</td>
<td>1.50</td>
</tr>
<tr>
<td>10</td>
<td>3.39</td>
<td>1.92</td>
</tr>
<tr>
<td>25</td>
<td>5.84</td>
<td>2.66</td>
</tr>
</tbody>
</table>

All other things equal, the tables show that **these items will increase the amplifying effect** of tail dependence on joint exceedance probabilities:
4.7.5  Implications

Correlation assumptions are an important determinant of the joint behaviour of risk factors. However, they are not necessarily the sole factor. Particularly where a model is used to determine probabilities associated with extreme events, it is important to consider tail dependence and the extent to which it should be taken account in the model. Tail dependence may be allowed for explicitly in the choice of model or, if assuming a model which does not exhibit tail dependence (e.g. a Gaussian copula), in the calibration of its parameters.

For a well-diversified company using a VaR measure, the scenarios representing the SCR will be a combination of moderate changes in individual risk factors rather than being dominated by more extreme changes in a small number of risk factors. Allowing for tail dependence may be less significant for such companies.

When modelling a large number of relatively homogeneous risk factors where there is expected to be a common underlying risk driver tail dependence is likely to be an important consideration. Some examples include equity returns in different developed markets, changes in corporate bond spreads for different credit quality steps, etc. The example of the Student’s t copula in different dimensions illustrates this concept.

4.8  Application of statistical techniques

4.8.1  Introduction

Statistical techniques are commonly used in the selection, fitting and validation of models for individual risk factors, particularly for some market risks where the volume of data available lends itself to the application of such techniques. At first sight, the use of statistical techniques in the context of a dependency structure may appear futile in the context of life assurance due to the “curse of dimensionality”. As the number of risk factors being modelled increases, the limited number of data points available becomes less dense in “risk space” resulting in a reduction in the information about joint behaviour which can be extracted from them (see section 6.5 of [ref. SSS], for example). Techniques for the fitting and validation of copulas have been used in non-life insurance and in other fields such as hydrology, biology and econometrics.

Whilst data constraints do naturally limit the extent to which firm conclusions may be drawn from such an analysis and additionally the selection and parameterisation of a copula will depend on judgement rather than purely a calibration exercise, we nevertheless felt it would be useful to include some examples of their application in this paper. Firstly, because such techniques are perhaps less well known amongst life assurance practitioners. Secondly, and more importantly, we believe that such techniques can be helpful in informing and validating the judgements made. For example, if a decision has been made to use a Gaussian copula to describe dependency, these techniques can be useful in:

(i) Deciding whether or not to make an allowance for tail dependence;
(ii) Forming a view on the appropriate amount of any such allowance.
The approach we have adopted to model fitting and testing is based on ideas presented at a Knowledge Transfer workshop led by Christian Genest and Johanna Nešlehová organised by the Scottish Financial Risk Academy in May 2012 [ref. GN] and on extensions of the Kolmogorov-Smirnov and Anderson-Darling tests introduced by Malvergne and Sornette [ref. MS] for the Gaussian copula and Kole, Koedijk and Verbeek [ref. KKV] for the Student’s t copula. The use of coefficients of finite tail dependence and conditional probabilities to select copula assumptions is based on work by Makin and Stevenson [ref. MaS].

In practice, we believe that graphical techniques which enable stakeholders to visualize the concepts and the judgements they are being asked to endorse are more effective. This is inevitably a challenge when dealing with multivariate models where 25 or more risk factors may be used, with the number increasing significantly depending on the number of business units included in the scope of the model.

In order to illustrate the techniques, we have used the following 3-dimensional data set.

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>Data series</th>
<th>Start Date</th>
<th>End Date</th>
</tr>
</thead>
</table>

All analysis has been based on monthly increase in each risk factor during the period 31.12.1996 to 31.12.2014 (i.e. 216 data points).

4.8.2 Statistical method - overview

Figure 4-16 illustrates the general statistical method for the selection and testing of a model.
If there is a specific feature of the world which one is considering whether to model (e.g. whether there is dependence between two risk factors), one may test whether that feature is present in data. These tests can be performed qualitatively (e.g. by visual examination of the data). Alternatively, there may be statistical tests for the feature which can be applied.

If the feature is present, one may select one or more models which exhibit that feature. For example, a family of elliptic copulas with non-zero correlations may be used if it is considered that there is evidence to support a relationship between changes in risk factors.

Next one selects the parameters for the model(s). This can be performed using quantitative techniques such as Method of Moments, Maximum Likelihood or, as we shall see for copulas, hybrid techniques. Alternatively, one may use techniques which are not solely driven by the data but instead use the data to help inform judgement. Such techniques may be appropriate, for example:

- If one has prior views about the future relationship informed by general reasoning or expert judgement, particularly if these are not fully reflected in the historical data;
- To reflect the uncertainty in the parameters;
- If a rounding convention has been adopted, perhaps to avoid spurious accuracy (e.g. if using correlations which are restricted to integer multiples of 10% or 25%, say).

Examples of these techniques are provided in appendix A.4.

Finally one may test the model using statistical goodness of fit tests. Where a family of nested models has been postulated where one model is nested within another by placing constraints on the parameters (e.g. a Student’s t-copula is nested within an IT copula by constraining all degrees of freedom parameters to be equal), one may apply further tests or criteria to assess whether use of a more complex model with additional parameters is warranted. For example, the Likelihood Ratio test provides a statistical test to establish whether a particular model within a nest family should be rejected in favour of a more complex version. Decision criteria based on Likelihood Functions with penalties for the number of parameters (e.g. the Akaike Information Criterion or Bayesian Information Criterion) can be used to inform decisions about whether an increase in the number of parameters is warranted. These techniques are described further in appendix A.6.

The process described above is not generally sequential. Results from fitting or testing a model may suggest an extension of the family of models being considered, for example. In particular, the boundary between fitting and validation of a copula model is not necessarily clear cut and it may be preferred to make maximum of use of
any relevant data to inform the parameterisation rather than “keep some back” for validation. In that event, the validation may rely on other more qualitative techniques such as general sense checks.

Where the dependency structure forms part of an Internal Model to be used to calculate the SCR for Solvency II reporting, firms must ensure that the model meets the relevant Statistical Quality Standards including those relating to choice of data, adequacy of actuarial techniques, expert judgement, documentation and the use test.

Where the model and parameters are being recommended by an actuary, the relevant Technical Actuarial Standards apply. TAS-M: Modelling is of particular relevance and is closely aligned with the Statistical Quality Standards of Solvency II. In particular, TAS-M requires that a model is fit for purpose and is no more complex than can be justified given that purpose. This is particularly important from the perspective of users of the model where transparency and understanding of the choices they are being asked to make and the implications of alternative choices are vital.

4.8.3 Statistical techniques – example

We have included examples of the application of the techniques described above to the data set described in section 4.8.1.

As one might have anticipated, statistical tests provide strong evidence that the three risk factors are not independent. However, goodness of fit tests do not lead to any firm conclusions regarding the copula model with the majority of the tests not leading to rejection of the Gaussian copula. Nevertheless, the results of model fitting, the p-values from the goodness of fit tests and the results of the Likelihood Ratio tests and Information Criteria do suggest the presence of tail dependence, particularly for EQ/CR. The results also suggest that the degrees of freedom parameter for a Student’s t-copula varies between pairs of risk factors.

Therefore, one is faced with the following choices:

- Should we use a Gaussian copula or model tail dependence explicitly?
- If we use a Gaussian copula, what allowance, if any, should we make for tail dependence?
- If we use a Student’s t copula, to what extent is the single parameter for degrees of freedom really a limitation? For example, is tail dependence really only a significant factor for a small number of pairs of risk factors so that we can focus on a suitable calibration for these and are less concerned about the financial impact of simplifying assumptions for other pairs?
- Would a hybrid model (e.g. a Grouped T model) be a suitable compromise?

4.8.4 Correlations and confidence intervals

The copulas typically in use within the UK life assurance industry at present have a correlation matrix as one of their inputs.

One graphical method that may be useful in the selection of correlation assumptions and showing the uncertainty around them uses charts of rank correlations over different time periods and confidence intervals around them. This provides a relatively compact visual representation which some stakeholders may find easier to digest than tables of numbers.

For example, one might produce a chart showing rank correlations:

(a) From a varying start date to a fixed end date (e.g. the end of the data period);
(b) From a fixed start date (e.g. the start of the data period) to a varying end date;
(c) Over a moving window.

Under (c), the length of the window could be selected using judgement as being sufficient to form a view on “short-term” correlations and give an idea of how correlations could vary between benign and stressed conditions.

Confidence intervals around the correlations could be produced using two techniques:
(i) Using analytic techniques based on the Fisher Z-transform. These use asymptotic properties of the distribution of transformed data to provide formulas for the upper and lower bounds of a confidence interval. There are various versions of the formula which are intended to adjust the result to allow for the finite sample size. See for example [ref. R].

(ii) Using bootstrapping techniques where a large number of synthetic data sets are generated by resampling the data with replacement. The rank correlation for each of the synthetic data sets is then calculated. This process generates a large number of simulated values of the rank correlation from which appropriate percentiles may be drawn to determine the confidence interval.

Figure 4-17 is an example of a chart of type (a) for our EQ/CR data set.

Figure 4-17 EQ/CR Spearman's Rank Correlation

Figure 4-17 shows the values of the Spearman rank correlation estimated from a varying start date to a fixed date of 31.12.2014 together with a 95% confidence interval. The values have been generated using the analytic approach (with no adjustments) and bootstrapping (with 1,000 simulations).

Note that this approach provides information only about the rank correlation – a scalar statistic which is “global” rather than relevant to a specific area of the joint distribution such as the tail. The above technique may be
helpful in informing one’s best estimate view of a correlation (e.g. by selecting an assumption which lay within
the confidence intervals over a long period of time).

Allowance for tail dependence is a different question. One crude approach would be use the confidence
intervals to inform an assumption which was believed to be prudent. A more refined approach using conditional
probabilities is described in the next section.

4.9   Model validation - backtesting using coefficient of finite tail dependence

4.9.1   Introduction

As described in section 4.7, one of the main challenges in selecting and parameterising a copula is the extent to
which it is appropriate to make allowance for tail dependence. As we saw in section 4.7.2, tail dependence is an
asymptotic measure of conditional probabilities in the tails. In section 4.7, we saw that the importance of tail
dependence was a function of the risk measure and the extent to which the “biting scenario” corresponded to
extreme or more moderate changes in the individual risk factors.

One possible approach to the assessment of the appropriateness of a given dependency model is to compare the
conditional probabilities for pairs of risk factors resulting from the assumed model (i.e. the coefficients of upper
and lower tail dependence – see below for a definition) with the values observed derived from the sample data
(i.e. the empirical coefficients of finite tail dependence).

The coefficients of finite upper and lower tail dependence are functions which explain how conditional
probabilities vary according to percentiles. In the limit, their values tend to the corresponding coefficient of tail
dependence. An overview of tail dependence together with a detailed explanation of the coefficient of tail
dependence and coefficient of finite tail dependence was provided in section 4.7.2.

This approach permits the use of graphics and may therefore be a useful tool in explaining to stakeholders the
choices available to them. It may also be used to select assumptions – this illustrates the blurred distinction
between assumptions selection and validation.

4.9.2   Empirical coefficients of finite tail dependence

The empirical coefficients of finite tail dependence functions are obtained by counting the number of pseudo-
observations in our sample that fall into the shaded rectangles in Figure 4-12. For example, if we have a sample
of N observations \((X_i, Y_i)\) \(i = 1, \ldots, N\) from \((X, Y)\) with ranks \((R_i, S_i)\) then the empirical lower tail dependence
function \(\hat{\lambda}_L(q)\) is given by the ratio:

\[
\hat{\lambda}_L(q) = \frac{\# \{ R_i/N \leq (1-q) \ AND \ S_i/N \leq (1-q) \}}{\# \{ R_i/N \leq (1-q) \}}
\]

where “\#” denotes the number of observations which satisfy the condition(s) inside the curly brackets. This is
simply the number of actual observations in our sample where the values of both risk factors are less than the \((1-q)^{th}\)
quantile divided by the number of observations of the first variable \((R)\) which are less than the \((1-q)^{th}\)
quantile.

The empirical upper tail dependence function can be constructed in a similar way by reversing the ranks of the
observations.

As an example, if \((X_k, Y_k)\) is a specific observation of \((\text{Equity Return, Increase in Credit Spread})\) with rank \((R_k, \ S_k)\) then the value of the lower tail dependence function at \(1-R_k/N\) is obtained as follows:

(a) Let A be the number of observations of the \((X_i, Y_i)\) whose ranks \((R_i, S_i)\) satisfy \(R_i \leq R_k \ AND \ S_i \leq S_k\).
(b) Let B be the number of observations of the \((X_i, Y_i)\) whose ranks \((R_i, S_i)\) satisfy \(R_i \leq R_k\).
(c) The value of the empirical lower tail dependence function is given by \(A/B\).
In practice, we can choose the labels \( i \) so that the \( R_i \) are ordered in non-decreasing rank (i.e. assuming there are no ties, \( R_1 = 1, R_2 = 2 \) etc.).

The algorithm above then simplifies so that the value of the lower tail dependence function at the point \( 1 - k/N \) is then obtained as follows:

(a) \( A = \text{number of observations where } i \leq k \text{ AND } S_i \leq S_k \).

(b) \( B = k \)

i.e.

\[
\hat{\lambda}_L \left( 1 - \frac{k}{N} \right) = \frac{\# \{ S_i \leq k, i = 1, \ldots, k \}}{k}
\]

4.9.3 Adjustments for negative correlations

Where there is a negative correlation between risk factors (e.g. equity returns and increases in credit spreads), it is generally the tail dependence in the upper left hand quadrant (i.e. a fall in equity values combined with an increase in credit spreads) that we are interested in. In order to provide meaningful values for tail dependence, it is therefore necessary to adjust the data by multiplying the sign of one of the random variables in each such pair by -1.

4.9.4 Use in validation

A visual assessment of goodness of fit can be performed by comparing the empirical coefficient of finite lower and upper tail dependence function \([\hat{\lambda}_U(q) \text{ and } \hat{\lambda}_L(q)]\) with the coefficient of finite lower and upper tail dependence \([\lambda_U(q) \text{ and } \lambda_L(q)]\) for the proposed model.

Confidence intervals around the theoretical tail dependence functions \(\lambda_U(q) \text{ and } \lambda_L(q)\) for the proposed model given a sample size \(N\) for the data can be estimated using bootstrapping techniques. Bootstrapping is a standard statistical technique and involves generating a large number of simulations of sample size \(N\) by sampling at random from the assumed model. Each simulated data set of size \(N\) represents a set of observations that we might have seen if the underlying model was “correct”. The corresponding values of the empirical coefficients can be determined in each simulation and percentiles estimated.

If the empirical functions lay outside the confidence intervals then that would provide evidence of a poor model fit. A model which appears to be a poor fit globally may nevertheless be considered adequate if it produces conditional probabilities which are consistent with the observed data in the desired part of the distribution (e.g. at a chosen point in the tail).

Note that for all the copulas we consider (Gaussian, T and IT) are radially symmetric so that the model coefficients of finite tail dependence are equal; i.e. \(\lambda_L(q) = \lambda_U(q)\) for all \(q\).

This is best-illustrated by an example.

Figure 4-18 below shows:

| Coefficient of finite tail dependence for a bivariate Gaussian copula with a correlation of 50% | Solid red line |
| Upper 97.5%-ile and lower 2.5%-ile confidence levels for the empirical coefficient of finite tail dependence for a sample size of 211 where the “true” underlying copula is a Gaussian copula with a correlation of 50% | Solid mid-blue lines |
| Empirical coefficients of upper and lower tail dependence derived from the Equity/Credit data. Note that the signs of increases in credit spreads has been multiplied by (-1) for the purposes of this chart. | Gold circles (negative equity returns, increases in credit spreads) Blue crosses (positive equity returns, decreases in credit spreads) |
Figure 4-18 EQ/CR Bivariate Gaussian Coefficient = 0.5

Figure 4-18 shows some asymmetry in the tail with the conditional probabilities associated with extreme upwards movements in credit spreads and falls in equity values somewhat higher than those corresponding to movements in the opposite direction.

The red line in Figure 4-18 tends towards zero. This is consistent with the Gaussian copula having a coefficient of tail dependence of zero.

The confidence intervals expand as events become more extreme, reflecting the decreasing volume of data in the tails.

Although the gold circles do not appear inconsistent with the confidence intervals, they are at the boundary, particularly around the 80th percentile. If a Gaussian copula is to be used, this may suggest use of a stronger correlation assumption if the SCR biting scenario includes percentiles of 80th or above in credit and equity.

Figure 4-19 below is similar but with the Gaussian copula model replaced by the bivariate T-copula estimated using MLE techniques.
Comparing the two charts, it is apparent how the T-copula lifts the conditional probabilities in the tail of the distribution (consistent with a non-zero coefficient of tail dependence). The red line also no longer converges to a value of zero.

The T-copula appears consistent with the data across a wider range – in particular, in the extreme tail and in the body of the distribution.

4.9.5 Application in assumption setting

The techniques described above may be used to assist in assumption setting. For example, if it has been decided to use a Gaussian copula model, then one may use a correlation assumption estimated from data using one of the techniques described in section 4.8.4 and appendix A.4 to inform one’s central view of an appropriate assumption. However, if there is evidence to support non-zero tail dependence, it may be desired to adjust that correlation to include an allowance for tail dependence.

This could be done:

(a) Visually – for example, adjusting the correlation assumed in the Gaussian copula moves the red line (and the confidence interval about it) up and down whilst the empirical conditional probabilities based on the sample data do not move. The correlation can be adjusted until the desired relationship between the modelled conditional probabilities and the empirical conditional probabilities is reached.
example, if one has a view on the corresponding percentiles in the SCR biting scenario, one may adjust the correlation of the Gaussian copula to obtain a model consistent with the empirical conditional probabilities at that percentile.

(b) Analytically – the visual technique above could be supplemented by comparing the values of the empirical conditional probabilities with those produced by the Gaussian model for various correlation assumptions at the desired percentiles.

(c) Targeting – the approach in (b) could be modified by requiring the modelled conditional probability to be equal to a specific value at a given percentile. That specific value could be selected on the basis of expert judgement or based on the data. This approach is suggested in Venter [ref. V2] for example.

By means of illustration, Figure 4-20 to Figure 4-23 below show how the coefficient of finite tail dependence changes as the correlation assumed in the Gaussian copula is strengthened. Conditional probabilities in the tail of the distribution are increased and the red line rises. If one had a view on the percentiles underlying the SCR biting scenario, one could aim to use a correlation assumption which produced values of the coefficient of finite tail dependence at that percentile which were consistent with a desired target level (e.g. a level consistent with empirical values or some other value judged to be appropriate). “Wiggling” the red line and its related confidence interval around in this way could help increase engagement by stakeholders in the assumption setting process.

Figure 4-20 EQ/CR Bivariate Gaussian Coefficient = 0.6
Figure 4.21 EQ/CR Bivariate Gaussian Coefficient = 0.7
Alternatively, if there was a desire to retain a model with explicit tail dependence, one could use such plots to inform the choice of parameters. For example, Figure 4-23 illustrates one potential choice of parameters for a T-copula which appears to more closely match the gold coloured tail.
4.9.6 Typical allowance for tail dependence

As discussed in section 1.1, the majority of UK life insurers that intend to use a copula based approach to calculate their Solvency II SCR, also intend to use a Gaussian copula, at least initially. Some other insurers are planning to use a covariance matrix approach for their economic capital models.

The 2015 Internal Model Calibration survey produced by Towers Watson indicates that 12 insurers were including an allowance for tail dependence in correlations involving market risks and provides a breakdown of that additional allowance as follows:

<table>
<thead>
<tr>
<th>Additional allowance for tail dependence</th>
<th>Number of companies</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;15%</td>
<td>4</td>
</tr>
<tr>
<td>10%-15%</td>
<td>2</td>
</tr>
<tr>
<td>5%-10%</td>
<td>3</td>
</tr>
</tbody>
</table>
4.9.7 Further refinements

The approach described above takes a relatively simple view of conditional probabilities by using the same percentile for each risk factor. One further refinement of this approach would be to allow the percentile to vary for both risk factors. For example, one could draw a grid by subdividing the square \([0, 1] \times [0, 1]\) into smaller square and consider the probabilities of observations falling into each square under the assumed copula model with the proportions observed in the data.

4.9.8 Assumptions where there is no/little relevant data

The techniques described above can be useful to support the selection of an assumption where there is a reasonable volume of relevant data. In practice, this means that they are useful only for certain market risks. For other types of risk, the selection of an assumption must necessarily be based on judgement. However, even here, the assumptions used for market risks can be informative when choosing any additional allowance for tail dependence, for example.

Items which may be considered include:

Causal relationships – does a change in one risk factor have a tendency to result in a change in the other? (and vice versa). How strong is that relationship?

Common underlying risk factors – is there a common factor that has a tendency to “drive” simultaneous changes in two or more of the modelled risk factors? Does it result in large changes happening together, suggesting the presence of tail dependence? (Recall that the Student’s t copula is an example of a “Normal mixing distribution” where a common factor scales values produced by a Gaussian copula, resulting in the introduction of tail dependence.)

Tail dependence – are there any reasons to believe that the relationship between changes in two risk factors becomes less diffuse in the tails resulting in a greater alignment between the ranks of changes in one risk factor and the other (e.g. due to the presence of a common underlying risk factor)?

Sensitivity – how important is the assumption in the context of the overall result? This will have implications for the level of detail to be considered and documented.

The above may assist in forming a view about the level of any correlation and the extent of any further allowance to be made for tail dependence. The amount of any additional allowance for tail dependence may also be informed by the choices made in calibrations for market risk factors.

As for all areas of expert judgement, it will be important to set out clearly the rationale for the assumptions and the sensitivity of the model output to the assumptions. An appropriate governance process should also be followed. It may be useful for the governance to involve a number of experts from different areas of the business to ensure that, overall, the assumptions are coherent. A rigorous treatment of expert judgement process can be found in [ref. EJWP].

4.10 Internal consistency (Positive semi-definiteness)

A problem faced by users of elliptical copulas is the need to ensure that the proposed correlation matrix is internally self-consistent and is a genuine correlation matrix. All correlation matrices satisfy an internal consistency condition known as “positive semi-definiteness” (PSD). Broadly speaking this means the correlations between all possible n-tuples of correlations are consistent. Thus, for example, if the correlation between risk pairs \((X,Y)\) and risk pairs \((Y,Z)\) are large and positive, one would expect the correlation between pair \((X,Z)\) to be large and positive – see for example [ref. H]. One would expect correlation to be a broadly transitive relationship. Mathematically, this means that the eigenvalues of the matrix must all be non-negative. Indeed, some copula simulation algorithms will only work if the matrix is positive-definite. This is a stronger condition and requires all the eigenvalues to be positive, or, equivalently that the matrix be PSD and all its columns are linearly independent. (Note that the covariance matrix approach is a formula and will produce a
result in most circumstances. However, if the underlying matrix is not PSD, there will be some combination of risk factor changes which result in zero or complex numbers being produced.)

Correlation matrices derived from a complete dataset where all risks are sampled contemporaneously will always be PSD by construction. However, as we have seen above, life insurers seldom have the luxury of having a set of data for all risks that were recorded over the same period of time at the same frequency. A candidate matrix may not be PSD due to inconsistencies introduced by the following:

- Selection of assumptions on a pair-wise basis using different periods of time and/or different sampling frequencies.
- Use of judgement in the selection of assumptions.
- Inclusion of additional allowances for tail dependence.
- Any rounding convention used when selecting correlations.
- The correlation matrix has been amended in order to perform sensitivity testing of correlation assumptions.

If the initial candidate matrix is not PSD, then it may require adjustment prior to use. There are several techniques available which vary in complexity. One of the more straightforward techniques involves elimination of negative eigenvalues. The candidate matrix is first diagonalised using standard techniques of linear algebra. Any negative or zero eigenvalues appearing along the diagonal are replaced by a small positive value chosen by the company and the resulting matrix transformed back.

Alternative techniques which seek to find the PSD matrix which is in some sense “nearest” to the initial candidate matrix are also available – see, for example, the work of Nick Higham and colleagues. The latter techniques seek to minimise the distance (measured according to some metric) between the original candidate matrix and the space of valid matrices.

As the resulting matrix will necessarily differ from the initial matrix, the process of selecting a PSD matrix involves certain choices. For example, one may choose to apply a weighting scheme within the metric which allows certain assumptions to vary to a greater extent than others (if one has stronger views about the latter set of assumptions). If one has such strong views about a particular sub-matrix of assumptions that one wishes to keep them fixed and allow only others to vary, then one can include this as a constraint in the optimisation process (assuming, of course, that the sub-matrix is itself PSD).

Whichever technique is used, users will need to be satisfied that the resulting matrix remains consistent with the views reflected in the selection of the original candidate matrix. One may do this by inspection or introduce quantitative acceptance criteria. For example, one may inspect a histogram of changes or require that a given proportion of changes fall within certain limits. However, this does not take into account the implications for capital requirements. One may be willing to accept a larger movement in the less material correlations that have a relatively small impact on capital requirements. One could therefore require that the effect of changes on capital requirements produced by a covariance matrix approach did not exceed a certain limit. (The covariance approach is used as it is not possible to determine capital requirements for the original candidate matrix using a copula based simulation approach).

If the adjusted matrix does not meet the specified acceptance criteria, one may then have to resort to a more sophisticated technique or choose a different set of adjustments. This could lead to several iterations. Failure to produce a suitable PSD matrix may indicate a more fundamental inconsistency within the original candidate matrix which needs to be reviewed or that a difference choice of copula may be appropriate.

4.11 Top-down validation tests

Having chosen a copula, selection of the parameters is generally a bottom-up process. However, it is important that consideration is given to whether the resulting assumptions are collectively reasonable and fit for purpose. There are various ways by which this may be assessed. A number of examples are provided below.

12 Nearest Correlation Matrix, Nick Higham (See [ref. H])
**Sensitivity testing**

Assumption sets underlying the dependency structures used in life company internal models typically have high dimensions. For example, if there are 25 risk factors, the correlation matrix underlying a Gaussian copula will have 300 distinct parameters. In testing assumptions, it is important to focus on those which have the most significant impact on the model output. These can be identified by testing sensitivities to individual assumptions. In general, a covariance matrix approach will be quite adequate for this purpose rather than re-running the full model many times.

When testing sensitivities, one should have regard to the relationships between changes in risk factors. For example, strengthening the correlation between risk factors X and Y may suggest strengthening other correlations involving X or Y for reasons of maintaining internal consistency. A sensitivity to a change in one correlation in isolation may not give a reasonable view of the total change if one were to make corresponding changes in all related assumptions.

**Peer review**

Proposed assumptions could be subject to review by one or more individuals with relevant expertise from around the business. Meetings could be held with the purpose to review and challenge proposals. These meetings could examine the rationale for particular assumptions, perhaps identifying relationships between changes in risk factors which had not been fully considered or whether conclusions drawn from analysis of data were reasonable or required further adjustment. The review should also consider whether the assumptions were appropriate on a prospective basis and were not unduly driven by historical data, taking into account the scarcity of data, the uncertainty in the analysis and expectations regarding the future relationship between changes in risk factors.

**Independent review**

Proposals could be reviewed by individuals who are independent from the formulation of the original proposals and subject to a different reporting line. The idea behind this is that the reviewers should be free of any influence from those responsible for development of the model, which, in theory, should lead to a more objective review.

The Internal Model requirements of Solvency II require an independent validation process. A typical process would be for the assumptions to be developed by one or more individuals in the Finance or Actuarial functions (perhaps using a panel process, as described above) with independent validation carried out by a 2nd line Risk function.

**Scenario analysis**

One could examine the simulated scenarios which generate losses close to the SCR and ask stakeholders for their view on whether these were reasonable given their knowledge of the business. One could examine scenarios which, when ranked by losses, lie in a “window” around the scenario corresponding to the SCR. Alternatively, one could average out the scenarios in the window, perhaps using some form of kernel smoothing, to determine a scenario which was in some sense representative of the scenarios giving rise to the SCR.

One could then ask stakeholders whether the scenario(s) corresponded to their expectations. For example, if a risk factor to which the business had a significant exposure featured relatively weakly in the scenario, this could prompt questions about whether the strength of the associations between that risk factors and others was appropriate. Alternatively, if a certain risk factor featured relatively strongly in the scenario(s), does this seem reasonable given exposure to this risk factor and expectations of stakeholders regarding the strength of its relationship with other risk factors?

Conversely, one could ask stakeholders to postulate a scenario involving changes in several risk factors, evaluate the “heavy” or proxy model on that scenario and determine its corresponding ranking or percentile in
the overall distribution of losses. One could then ask stakeholders to form a view on whether the ranking of the scenario seemed reasonable.

In practice, it may be very challenging to ask stakeholders to assign a probability to losses under a particular scenario. However, presenting scenarios can make the calculations feel more “real”, help engage stakeholders in discussing the relationships between risks and provide a high level sense check on the results.

**Industry benchmarking**

Several of actuarial consultancies produce annual surveys comparing practices, models and calibrations. Comparisons based on survey results will necessarily be subject to limitations. Companies do not all adopt a common definition of a risk factor, for example. Companies with low exposures to a given risk factor may be less averse to adopting a relatively strong assumption for correlations between that risk factor and other risk factors. Copula models can differ. A correlation assumption of a company using a Student’s t copula is therefore not necessarily directly comparable with that of a company using a Gaussian copula, particular if the latter chooses to include an allowance for tail dependence.

Nevertheless, comparisons based on surveys can be useful in highlighting assumptions which appear out of line and help focus attention on the rationale for those assumptions.

**4.12 Final comments**

In this section, we described some of the techniques which can be used to select, parameterise and test a copula.

In practice, the most important factor in selecting a copula appears to be that of transparency and the use test. The mathematics of copulas is complex and it is not impossible that as actuaries and users of the models become more familiar with the techniques, greater sophistication may be employed.

Graphical tools and simple numerical examples are useful in explaining concepts to stakeholders rather than mathematical formulae. The disadvantage of graphical tools is that they are most effective in two dimensions. For this reason, it appears to us that vine copulas may be a fruitful line of future research.

One of the most important decisions to be made in selecting and parameterising a copula is how to make an appropriate allowance for tail dependence. We have illustrated one possible tool based on conditional probabilities which makes use of graphical techniques to inform the choice of assumptions.

Ultimately, there is a significant level of judgement involved in selecting and parameterising a copula. The resulting assumptions should be subjected to overall reasonableness tests (e.g. consideration of the scenarios giving rise to losses similar in magnitude to the SCR, sensitivity testing) to ensure that they are plausible, given the risks to which the business is exposed.
5 Proxy Models on Copula and Proxy Modelling Method

5.1 Overview

In previous sections of this paper we have outlined the drivers for developing detailed capital models, and in particular those that use a copula-based simulation approach. This enables firms to generate a rich and complete distribution of profit and loss, which can inform regulatory and economic capital requirements, and assist firms to understand the nature of the risks that they face and their impact on the balance sheet. However, such simulation techniques are only useful if we can, with reasonable accuracy, calculate the financial impacts on an insurer’s balance sheet over a wide and rich set of points. For example, for a credible distribution of profit and loss, firms would expect to run hundreds of thousands, or even millions of samples from the copula distribution. In most cases, current technology prohibits firms from running heavy models at all of the required simulation points. Because of this, many firms have developed ‘proxy’ or ‘light’ models, which are used to estimate the relevant financial metrics, such as movement in surplus. This topic, and in particular the technical grounding, was discussed in some detail by Hursey et al (see [ref. HM]). This section attempts to address more practical issues with developing, validating and communicating proxy models.

In this section we provide a discussion of all the considerations that firms should consider when implementing a proxy model in order to meet its objectives. We also consider the validation of these models, in particular how to interpret and communicate the validation results in order for senior management to make decisions about the use of the model for different purposes.

5.2 Types of calculation model

In this section we briefly outline the types of calculation model that could be adopted to support a copula approach.

5.2.1 Traditional Actuarial Models

Traditional actuarial models, now often referred to as ‘heavy models’, typically aim to project liability cashflows and derive present values. Usually these models project on an individual policy basis, or sometimes with grouped ‘model points’ as long as the grouping effect is not detrimental to accuracy. These models aim to project the cash flows of an insurance policy to a high level of accuracy, incorporating all the features of a product and the company’s expenses. Such models are inherently time consuming to run, even once for a ‘best-estimate’ calculation. Even with the computing power we have today companies will face challenges when trying to use these heavy models to calculate the balance sheet at thousands or millions of scenarios as required for a full and risk Probability Distribution Forecast.

For this reason, heavy models are unlikely to be a suitable partner for a copula approach. There may be some cases where they are appropriate:

- If we are looking to produce a less rich loss distribution, perhaps one that contains hundreds rather than millions of simulations. In some circumstances this may be appropriate, but such a loss distribution is unlikely to be valid for capital purposes, for example, because of insufficient richness in the tails of the distribution.
- Smaller companies, with relatively few and simple product lines may be able to produce calculations very quickly, so potentially support thousands of valuations in an acceptable timeframe.
- In some cases, the risk factors within the copula may be limited to ones that lend themselves more easily to a heavy model calculation. For example, if an annuity provider only models its interest rate and spread risk using a copula – perhaps as part of a Partial Internal Model under Solvency II – then the liability cash-flows can easily be revalued using different discount rates without needing repeated re-projections of the annuity payments and expenses.

In general though firms will usually struggle to use detailed heavy models to support a copula approach.

5.2.2 Accelerated / Streamlined Actuarial Models

An alternative would be to attempt to speed up the heavy actuarial models themselves to permit a large number of simulated balance sheets to be calculated, for example:
- Harnessing a large enough amount of computing power, for example via cloud computing to gain access to a massive number of processors. This technology has been used in other industries to manage and process very large volumes of data.

- Streamlining calculation models may also speed up calculations. However, the extent to which this is possible, given that most firms would have already optimised the calculations to some extent, will likely be low without invoking approximations.

- In the past many firms have adopted a data modelling approach to facilitate quicker calculations (e.g. using grouping of data into 'model points' or other forms of data compression). This would ultimately mean grouping policies together, itself creating a certain amount of approximation depending on the severity of the grouping.

This problem is not new. In the UK since the early 2000’s, companies with with-profits funds have been required to implement measures to enable the stochastic valuation of options and guarantees in order to produce realistic balance sheets. Most firms in this case looked to compress policy data through grouped model points in order to produce 1000 to 5000 simulated balance sheets. However, in this case, the stochastic variables were generally limited to key market variables, and other assumptions were modelled deterministically, making the use of a heavy model much easier.

Even if we are able to streamline the model, or acquire enough computing power, to enable thousands or millions of calculations to be performed, the firm would still have to set up basis files for each of those valuations. It is true that this too could be well automated – a computer could easily translate a risk factor outcome from the distribution into a change in valuation assumption and produce the updated tables – however, validation of model inputs and outputs would prove a very difficult task.

In the latter two methods, firms would need to consider whether the approximation introduced is actually greater than that implied with a proxy modelling method.

5.2.3 Proxy Models

The above issues would direct most firms towards adopting some kind of proxy modelling approach. These techniques have the aim of being able to estimate financial values connected to insurance liabilities without needing to perform a full model run. Most commonly, this takes the form of implementing either Proxy Functions or Replicating Portfolios. Whilst these two methods are different in their construct, they both adopt the same principle of trying to represent movements in the insurer's balance sheet in a simplified formulaic manner. In fact, Replicating Portfolios could be seen as a special case of a proxy function, as ultimately with this approach we are trying to derive an analytical formula to estimate asset and liability values.

Proxy Models, and in particular Proxy Functions, are the focus of the remainder of this chapter.

5.3 Communication challenges

In the past, senior management at insurance firms will have been aware that actuaries have developed and maintained large calculation models for valuing insurance liabilities. Governance frameworks and standards will have been established to demonstrate to the Board that the models are appropriate for use in producing financial statements and solvency assessments, for pricing new products, and for forecasting revenue and capital. These models will have been through thorough testing to validate that the models produce results that are consistent with the features of insurance products.

The use of proxy models provides challenges for communicating with and gaining the trust of senior management because:
They introduce an additional layer of approximation into stated results. The actuary can no longer state that the financial result obtained is the result of a detailed model calculation.

Communicating that a result from the proxy model is a ‘best-estimate’ assessment of the true result is difficult without some presentation of the potential error around that result. However the need to communicate uncertainty is not unique to proxy models and is also necessary for the heavy model runs themselves which rely on uncertain parameters.

Proxy models introduce an element of statistics and probability into the results. Insurance boards will be familiar with the nature of actuarial science, but to date this will have applied when setting the assumptions used for valuations, not within the model itself.

Firms are likely to look to use their proxy models to support a wide range of calculations and processes – for example economic capital, regulatory capital, balance sheet estimation, business forecasting and pricing. Whereas a heavy model is designed to project the true cashflows on the firm’s policies, and as such should be applicable for any purpose, the proxy model may be more or less appropriate for each of these uses. Senior management therefore have more uncertainty about the appropriateness of the proxy model for different uses.

5.4 Choice of Proxy Techniques

This section sets out a description of the two main forms of proxy model – Proxy Functions and Replicating Portfolios. We set out the advantages and disadvantages of each approach, and the relative usage of each method in the industry today.

5.4.1 Proxy Functions

This is a general technique wherein the user runs a number of fitting scenarios and performs a regression, so that the determinant variable (e.g. profit/loss, asset/liability value) can be expressed as a function of one or more explanatory variables (commonly referred to as risk factors). Hursey et al (2014) describe proxy functions in some detail, and a summary of the method and its advantages and disadvantages is discussed below.

5.4.1.1 Method

The broad steps required are:

1. **Define the determinant variable.** This is the variable that we are trying to estimate through the proxy functions. This could be the value of the asset or liability, the movement thereof, or a value that is derived from asset/liability values, such as surplus or free-capital. If the model is being used to calculate a value under a specific regime, then the determinant variable is likely to reflect that – for example Own Funds under Solvency II.

2. **Choose the potential explanatory variables.** Decide which risk factors should be included. This should be based on a fundamental analysis of the firm’s assets and liabilities and hence risks. Assessing which assumptions are feeding into valuation models is another way of ensuring that all the risks are captured. We should also be able to allocate certain risk factors to certain assets and liabilities for fitting purposes.

3. **Determine the process for designing and fitting the proxy functions.** There are several choices to be made when designing and fitting the model, and these choices will depend on the intended purpose of the model and idiosyncratic features of the firm in question. This is covered in section 5.5.

4. **Carry out fitting runs.** Set and carry out the heavy model runs required to generate enough points to fit the proxy functions. This would include a combination of single risk, or univariate, stresses, and probably some multivariate stresses which model more than one risk event occurring simultaneously. Note that the fitting runs can be carried out using a simple curve fitting methodology, or a more advanced methodology such as Least Squares Monte Carlo (see section 5.6).

5. **Fit the proxy functions.** Undertake the fitting of the proxy functions using a chosen regression technique. Different formulaic structures can be examined using either a mathematical approach or manual judgement. See section 5.7.

6. **Validate the proxy functions.** The goodness of fit of the proxy functions needs to be assessed through a number of means, including statistical analysis of both the in-sample fitting points, and further ‘out-of-sample’ points. See section 5.8.
The following sections looks at the advantages and disadvantages of using proxy functions.

5.4.1.2 Advantages of proxy functions

- The method is intuitive, and in most cases straightforward to explain. Univariate or bivariate functions can be shown graphically, which aids understanding for both the modeller and senior management.
- In some cases only a small number of runs are required to fit the curves adequately, for example when an exposure is intuitively linear we only need one fitting run (alongside the best-estimate), for quadratic curves only two points etc. Firms will run more than this to aid validation along the full curve, but usually only a small number (less than 10 perhaps).
- Running and fitting need not be computationally complex or onerous.
- May facilitate an automated fitting algorithm – which can assess a large range of formula structures and choose the most appropriate based on information criteria - which would be quick and low on manual effort and judgement.

The Least Squares Monte Carlo method can be used where assets and liabilities require stochastic calculations.

5.4.1.3 Disadvantages of curve fitting

- Can require high amounts of expert judgement when determining the form of the curves. Mathematical fitting algorithms remove some of the judgement in fitting, but themselves require some judgement as to what the parameters of the algorithm should be.
- May not be possible to fit functions to some complex assets and liabilities – for example where there is optionality - especially if technology restricts us to certain types of function e.g. polynomial to nth degree.
- Can be complex for large insurers with a large range of assets and liabilities.
- Can be prone to significant fitting errors in some cases, for where there are complex interactions between risks.

5.4.2 Replicating Portfolios

This technique aims to use portfolios of assets to replicate the value of the firm’s liabilities and how they move in relation to various risk factors. The assets need not be real in the sense that they are traded in real markets i.e. they could be synthetic. If the replicating assets can be valued analytically, then the liabilities can be easily and quickly valued under a wide range of scenarios.

Replicating portfolios are particularly useful when the liabilities have embedded options and guarantees. These can be time consuming to model, as they would normally require a stochastic simulation of the liability cashflows under different investment returns. However, in some cases the liability profile could be adequately replicated by a portfolio of assets, including derivatives, which may have an easily computed closed form value. The chosen copula simulation model can be combined with the reference replicating portfolio to calculate the balance sheet in any required scenario.

5.4.2.1 Advantages of Replicating Portfolios

- Calculations can be very fast due to the nature of some replicating assets having closed-form formulae.
- Calculations can be performed under a large range of assumptions, including complex combinations, in quick time.
- Due to the known formulaic structure, it can be more accurate at capturing the impact of combinations of risks.

5.4.2.2 Disadvantages of Replicating Portfolios

- Can be theoretically and practically challenging to replicate all types of policy liability, even with a large range of candidate assets.
- Certain risk sensitivities are difficult to represent with assets, e.g. insurance risks, and business risks, so at best a replicating portfolio model can only be part of the solution.
- It is difficult to allow for dynamic features of liabilities e.g. policyholder or management actions.
- Does not handle liabilities with risks that have substantial interaction with market/credit risk, for example lapse risk on investment policies or longevity risk on annuities.
- It is difficult to perform a calibration of a replicating portfolio without significant expert judgement.

5.4.3 Relative usage of each method

The use of replicating portfolios in the UK picked up following the introduction of the Realistic Balance Sheet in the early 2000’s, however its use was never widespread. More recently, firms have focussed more on the development of curve fitting techniques, including Least-Squares Monte-Carlo (LSMC) which is discussed in more detail in section 5.6.2. A survey in 2014 found that only 2 firms (out of 32 surveyed) were using replicating portfolios to assess market risks. In contrast, the use of proxy models has seen an increase. The same survey showed that 26% of the surveyed firms were using curve fitting approaches to fit proxy functions in respect of market risk, and 23% for insurance risks. Given that 53% of respondents stated that they are intending to use the standard formula under Solvency II, we would expect that for firms that are applying for internal models, the percentage would be much higher.

For this reason, the remainder of this section focusses on the curve fitting approach to using proxy functions. However, much of the content around designing and validation of the model could equally apply to the replicating portfolio technique. Certainly, the communication issues identified in section 5.3 are equally applicable to either approach.

5.5 Proxy Model Design

Although proxy functions could be viewed as a simplistic tool, in reality firms need to design their calculation software to fit with its reporting needs and the specific structure of the organisation. A proxy model that is well designed will be able to produce all the outputs that a firm needs, for example economic capital can be calculated at various levels of the organisation’s hierarchy, and the capital can be allocated lower down if required to individual products or groups of products. The diagram below illustrates a generic model structure. It shows different levels at which proxy functions could be established, and at each level the firm could calculate profits or losses that therefore capture diversification between risks at that level.

The design of the model will be different between firms, and even within the same firm there could be more than one design to fit different purposes. Ideally however, the firm would establish a single model that fits all of its reporting purposes.

---

5.5.1 Design Considerations

The following are the sorts of considerations a firm should make:

5.5.1.1 What metrics does the model need to calculate?

The metrics that a firm needs to report on, both internally and externally, will affect the design of the model. The regulatory results that we are looking to calculate from the model will have an influence. The SCR will likely be a metric, or an approximate roll-forward calculation so that we can demonstrate continuous solvency monitoring. Or a combination of these. This can impact the hierarchical design of the model, and the level of granularity required.

There will also be internal reporting requirements, for example VaRs at certain confidence levels for risk appetite monitoring, or deep tail stresses for capital buffer setting. Again, this may impact the granularity of the model if we are looking to assess results for different products or legal entities.

Does the company need to report on the assets and liabilities separately, or can it combine them in some or all circumstances? Separate asset reporting may be instructive, for example in measuring and managing the risk of a credit portfolio. For many purposes we may be able to combine assets and liabilities, to the extent that we are able to fit curves adequately.

5.5.1.2 What is the structure of the company we are reporting on?

The company structure will have a large influence on the design. Firms will usually need to report results externally at legal entity level, and for designated Insurance Groups. However, firms may also want to report results at other internal layers, for example by geography or for certain holding companies.

Firms may also be required to perform calculations at sub-fund level. For example, under Solvency II, most with-profits funds and all matching adjustment portfolios are treated as if they were ring-fenced portfolios, where gains and losses cannot be shared with other parts of the business. The aggregation model must then be set up to be able to calculate results for those separate portfolios.

Finally, firms may have defined business units or product groups at which they manage the business commercially. This may necessitate output at that level of granularity. The more granularity however, the more curves that will need to be fit. Whilst granularity is required to the extent that homogenous assets and liabilities will be easier to fit adequately, too much will require excessive fitting and validation. There may be more approximate ways to allocate results down to more granular products, for example defining easily accessible risk drivers to approximate the exposure to risk.
Fitting additional metrics does not necessarily require additional heavy model runs, because the heavy model runs will already provide the granular output. However fitting additional risk factors likely will require additional heavy model runs, especially if we predominantly use univariate stresses.

5.5.1.3 What risk factors should be included in the model and at what level of granularity?

In general, the decision on which risk factors should be included is not within the scope of this paper and will be specific to individual firms. However there are structural considerations that need to be considered.

A firm may decide that the risk factors that are calibrated i.e. have probability distributions fit for them, are too detailed for the aggregation model. Ultimately, a risk factor only needs to be included to the extent that the firm wants or needs to assess the diversification with other risks. Too many risk factors mean too many coefficients need to be fit in the proxy functions, and, more correlations assumptions need to be made.

As an example, a firm may investigate and calibrate risk functions for persistency risk for each of its current and legacy products. This may be required to best represent the risk inherent in each of these portfolios. However, the firm may have little data or knowledge on how persistency experience in one group will affect the experience on another i.e. it has little on which to base a correlation between the two. In this case, for the purposes of the aggregation model, the firm may choose to combine some of the product groups, for example into broad product categories or into current versus legacy business. This would imply that the aggregation model assumes 100% correlation between the risks that have been grouped together – the firm would need to make a judgement on whether that is desirable, and whether reporting the grouped risk factor provides enough management information of which to manage the business.

This is sometimes described in terms of risk drivers and risk factors. Diversification effects are only captured between different risk drivers. However there may be multiple risk factors for a particular risk driver reflecting the different magnitude of the effects of the risk driver on the more granular categorisation of risks. Risk factors will have different probability distributions but may be perfectly correlated with one another.

Risk factors allow for greater granularity within the model without requiring additional calibration or aggregation. An example is credit risk for investment-grade and sub-investment-grade rated corporate bonds. There may be a single credit risk driver with a particular probability distribution. However a multiplier is applied depending on the credit rating of a bond.

5.5.1.4 What other features do we need to account for?

There will be other, more idiosyncratic, features of a firm that need to be reflected in the proxy model. These will include tax, fungibility restrictions, reinsurance, and complex risks. The challenges caused by these idiosyncratic features are described below, though this is not an exhaustive list:

Tax

- The need to calculate capital results net of tax. There are options around how to do this, for example whether we want to fit net of tax loss functions (pre-proxy model), or calculate the tax impacts within the simulation model (in proxy model), or whether we want to do this for certain results outside the model (post proxy model). It is likely that for the best accuracy, the tax impact would need to be assessed for each simulation within the model, as this would take account of the nature of the scenario and profit/loss realised, and any restrictions on the amount of tax loss absorbency that could be realised i.e. due to current levels of deferred tax liability.

Fungibility Restrictions

- Ring-fenced funds or other fungibility restrictions. Many firms will be subject to restrictions on the flow of capital across the business. Examples include:
  - With-Profits funds, where surplus assets belong at least partly to policyholders and are as such not available to cover losses outside that fund.
  - Separate legal entities within a group will likely have to cover solo capital requirements, thus restricting any ability to achieve diversification across a group.
Under Solvency II, Matching Adjustment portfolios are notionally ring-fenced and thus surplus in these portfolios cannot be used to cover losses elsewhere in the organisation. These restrictions may prevent firms from recognising the full diversification effect from their risk portfolios. Firms must choose between modelling these features within their proxy model, or allow for them separately.

Reinsurance

- Firms who utilise reinsurance arrangements will also want to take account of this in their models. The requirements here will depend on whether the firm needs to report results gross and/or net of reinsurance. In general, Solvency II reporting requirements require liabilities to be shown gross of reinsurance, with a separate 'reinsurance recoverable asset' on the other side of the balance sheet. Firms may also want to show capital results before and after reinsurance, for example to investigate the effectiveness of an arrangement. The same may also be true of other risk management strategies such as hedging of investment risk.

Complex risks

- For some risks, proxy function approaches are difficult to use, for example because the risk factor has no easily measurable 'value', or where the exposure to this risk is likely to be too complex to model using polynomial functions. A common example is operational risk, which is usually a scenario-based type of risk with no obvious measurable risk factor, and the nature of the risk is such that it is dominated by catastrophic events. This is difficult to capture using a proxy function approach. In such cases firms could choose to model that risk separately and either add it on to the wider capital requirements, or integrate it into the model. The latter would normally require the marginal distribution of losses to be included in the model in empirical form i.e. as a rich data set giving losses at different levels of probability, rather than by fitting a proxy function. In this case the marginal distribution would be Uniform, with the loss functions as the empirical losses at different percentiles.

5.5.2 Constructing the Model

The above considerations will influence the design of the calculation model, which will include:

- The bottom level functions that are fit i.e. the lowest level at which we can perform the calculations (for example product groups, legal entities).
- The hierarchical structure of the model, which will determine at what levels we may be able to simulate and order losses and thus calculate diversified results.
- Any constraints to the diversification that can be achieved e.g. due to ring-fencing restrictions.
- Where and how the tax impacts on profits and losses can be calculated.
- The set of risk factors that will be used in the model, and which products/entities these will be applied to.

Where firms are using third party software supplied by an external vendor, there may be constraints imposed by that software, for example a limited range of mathematical functions (including statistical distributions and copulas) that can be used.

Note that the above design decisions have not yet considered the methodology for fitting proxy functions. This is included in the next section.

5.6 Model Fitting

5.6.1 Objectives for model fitting

Section 5.5 described how the proxy model will be designed to meet the various objectives of the firm. This section addresses the choices around how to fit the model. Firstly, we need to define the objectives of the fitting process, and our suggestions for this are below:
• **Accurate:** The model must be well fit such that it fits well to a firm’s 'heavy' models over an appropriately wide range of scenarios. This may focus on achieving strong fit at particular points or scenarios, for example around the 99.5th percentile VaR to meet solvency capital requirements, or at other points as used by the business for risk measurement. However, it is a requirement of Solvency II that firms produce a full Probability Distribution Forecast so it will be necessary to achieve a strong fit across a range of quantiles. Testing of the accuracy requirement is part of the model validation process, as discussed in section 5.7.

• **Parsimonious:** The model should not be more complicated that is necessary. The fitting approach should not lead us to implement functions that are longer and/or more complex than is required. Higher order polynomial coefficients and joint terms should only be included if they materially improve the fit of the model. Testing of the parsimony objective is inherent in some mathematical fitting approaches – for example Akaike’s or Bayes’s Information criterion (section 5.7.3). Some firms use a bespoke information criterion to select the model. If firms are using expert judgement to determine the structural form of the proxy functions, it would be more difficult to prove the satisfaction of this objective.

• **Avoid over-fitting:** This would normally be defined as an approach that overly focuses on fitting to the sample observations, which in some applications means fitting to random noise within the sample. This is less likely to be relevant in insurance applications, where the asset and liability calculations are either deterministic or based on stochastic models with sufficiently many simulations to ensure convergence. However, over-fitting in the context of proxy models would mean placing too much emphasis on achieving an accurate fit in the areas that the fitting runs are performed. For example, over-fitting can lead to a more accurate fit in the body of the distribution, but results in the tails of the distribution that are not sensible due to turning points in the polynomial function that occur outside the fitting range. The range of scenarios over which the proxy model is valid should be clearly specified as a limitation and a trigger framework produced so that the curve fitting process is repeated where necessary. Graphical validation can provide a quick and instructive view of the curves that are fit (at least up to 3 dimensions) so that the behaviour can be sense checked.

• **Practical:** The fitting also needs to lead to practical proxy functions i.e. one that can be incorporated into the firm's simulation model. This may preclude certain functional forms or polynomial orders.

### 5.6.2 Methods to Collect Heavy Model Output

As discussed in section 5.4.3, there are two common methods of fitting proxy functions: Curve fitting and Least Squares Monte Carlo (LSMC).

Curve fitting refers to a relatively simple process whereby a firm uses its heavy models to produce a small number of stress points, and the firm then uses a regression method – Ordinary Least Squares or Weighted Least Squares to fit a curve. The heavy model runs will likely include a combination of univariate and multivariate stresses. The fitting process may or may not include a mathematical approach to determining the optimal functional form of the model.

LSMC is a method commonly used where firms have complex and path dependent liabilities, for example in with-profits funds. In these cases, the options and guarantees inherent in the policy contracts, coupled with the management actions normally modelled, means that stochastic models are normally used in producing the balance sheet valuation of certain assets and liabilities. Fitting proxy functions for an aggregation model requires the firm to consider a large range of stress scenarios to the balance sheet, which would necessitate extremely large numbers of runs to be produced. We encounter a nested simulations problem, because we are required to value the balance sheet a large number of times, and each valuation requires a large number of simulations. The LSMC model was designed to overcome this challenge has been implemented in many of these cases. In this method, a large number of 'outer simulations' -- say 50,000 evaluations of the balance sheet - is produced, each of which represents a combination of risk factor outcomes as required in the aggregation model. For each outer simulation, the stochastic models required to value certain items on the balance sheet are run with a small number (as little as two) of stochastic simulations, rather than the usual 1000 or above. For a particular outer simulation, the balance sheet will contain significant random sampling error because the stochastic models will not have converged as there are only a small number of simulations used. However, if a
large enough number of outer simulations are performed, then the 99.5\textsuperscript{th} percentile and other percentiles will be convergent as whole. In general the percentiles in the tails such as the 99.5\textsuperscript{th} percentile will be slower to converge than percentiles in the body of the distribution such as the median.

![Figure 5-2 Convergence with Large Number of Scenarios](image)

LSMC was discussed as an approach in some detail in Hursey et al (2014). It is not within the scope of this paper to discuss this in more detail. However, it is worth recognising that whilst the process for simulating results from the heavy models is very different for curve fitting and LSMC, the processes following this to fit and validate the curve are similar, and the principals the same. Ultimately, both approaches aim to develop a proxy function, which will usually be a multivariate polynomial. From this point on, curve fitting is used generically across both methods.

The following sections outlined the considerations firms should go through before embarking on a curve fitting exercise.

5.7 Curve fitting – considerations

5.7.1 General approaches to curve fitting

Whilst the coefficients are a result of the proxy function fitting, the functional form itself is influenced by the approach taken to curve fitting. By functional form we mean:

- The order of the polynomial for univariate terms
- The inclusion of joint-terms – bivariate terms or higher if required – and the order of polynomial for these.

The resultant functional form will depend highly on the approach taken to curve fitting.

Top-down vs. bottom-up

Approaches can be categorised according to whether they are top-down or bottom-up and whether they are theory-driven or data-driven. Top-down approaches start with functional forms (e.g. polynomials) with a large number of coefficients covering a wide range of possible complex relationships between the risk metrics and the risk drivers. Coefficients are successively set to zero according to the curve fitting method until the chosen proxy function is selected. Bottom-up approaches on the other hand start with simple functional forms (e.g. a linear relationship covering a single risk driver) and introduce additional terms to the proxy function. Again terms are added until the chosen proxy function is selected.

Theory-driven vs. data-driven

Approaches can be further categorised according to whether the selection method relies on the data produced from the heavy model runs alone, or on business intuition or economic theory as to how the relationship between the risk metrics and the risk drivers should behave. In practice firms may use a combination of these two approaches.
Under the top-down approach, there will be a need to determine the universe of possible coefficient terms. This will be the starting form of the proxy function and the maximum order of polynomial terms chosen and joint terms chosen may rely on theory rather than data. Under the bottom-up approach, certain relationships could be included in the initial simple proxy function, if there are business or theoretical reasons that they should be included.

For a given form of the proxy functions, say \( f(x, y, z) = a.x^2 + bxy^3 + cz \), the coefficients may be selected using a data-driven statistical method such as maximum likelihood estimation or minimising a particular loss function. Alternatively coefficients may be driven by theoretical or intuitive relationships. For example \( c = 1 \).

In order to choose between forms of the proxy functions, expert judgement may be used or a mathematical approach. The mathematical approach will likely be similar to the information criterion approach of AIC or BIC wherein the loss function is minimised subject to a penalty for including large numbers of coefficients.

**Manual vs. automated**

In principle a data-driven bottom-up approach could be fully automated, with the full range of scalar multiple of a single risk driver considered at the outset. This would be an exhaustive search with a fully specified loss function and number of coefficients penalty.

### 5.7.1.1 Top-down/theory-driven approach

A top-down approach would mean using the firm’s knowledge about its assets and liabilities, and its exposure to risk factors to determine the functional form of the model, or at least to restrict the number of possible coefficient of the model through expert judgement. There may be an economic theory or business intuition that the impact of a particular risk factor on a particular metric will be linear. This can be seen as a ‘theory-driven’ approach, as firms will be using theory as applied to their business to determine what the proxy functions should look like.

A firm would complete a fundamental analysis of its assets and liabilities in order to:

- Assign the relevant risk factors to different asset and liability types
- Determine the maximum order e.g. \( x^3 \) we should consider when fitting the polynomial. In some cases, the firm can use judgement to restrict some functions to linear form if they believe that the exposure to the risk factor is linear (e.g. exposure of unit-linked liabilities to equity prices), or if the quadratic term of the exposure is deemed immaterial.
- Determine the range of joint-terms to be included. A joint term is relevant where there is a financial dependence of the sensitivity to one risk factor on the movement of another. For example, unit-linked liabilities have a joint dependence on both equity prices and lapse rates. This example can be extended to three or more joint terms. The firm can uses its expert knowledge to determine which terms to include in the model. The validation will need to cover the implicit zero magnitude terms that are not included.

The advantages or a top-down/theory-driven approach are

- There can be a clear link to the theoretical exposure to risk, which aids understanding and communication of the fitted functions.
- The processing time for the heavy models is limited to those we think are relevant.
- The approach can allow for the bottom-up approach to be used where there are areas of uncertainty. For example where a firm is unsure of how significant a joint term may be. In such cases mathematical fitting algorithms can assist in understanding the problem.
- Documenting the judgements made provides good evidence of the understanding of the firm’s risks.
- There is no need to develop an automated algorithm.

The disadvantages of a top-down approach are
- Reliance on expert judgement which could result in aspects of the functional form being overlooked e.g. choice of joint terms. In such cases, repeated iterations which gradually introduce additional terms into the proxy function may be required. This may be a slow process, so is only suitable if a firm has a long enough development timelines before the model is required to be used.
- This iterative process may broadly replicate a bottom-up approach, so the top-down approach may be a more manual and time-consuming process.

5.7.1.2 Bottom-up/data-driven approach

An approach at the other extreme would be to run a very large range of stresses in the heavy models, both univariate and multivariate points, and use mathematical methods to reduce the proxy functions down to a suitable form. This approach would inherently start with a very large number of coefficients, and the aim of the mathematical approach would be to gradually reduce the proxy functions whilst maintaining an adequate level of fit. This could equally be done manually, but that could be very time consuming, so a firm may at least start with a mathematical approach. The possible approaches in general involve stepping through different functional forms, gradually reducing the number of coefficients, and testing whether that change maintains the quality of the fit i.e. that term removed was not adding sufficient ‘information’ or explanatory power.

The advantages of a bottom-up approach are:
- No assumptions have to be made about the limit on the possible highest order of polynomial terms that can be included. So there is lower chance that something important has been overlooked.
- Both approaches can work alongside a mathematical fitting algorithm, which can be easily automated and does not rely on manual intervention of expert judgement.
- Both approaches are consistent with a rigorous statistical approach which will demonstrate that the functional form enables adequate fit and is parsimonious.

The disadvantages of a bottom-up approach are:
- Can be prone to over-fitting to the heavy-model simulation data e.g. in areas where the fitting data is not rich it can lead to fitting terms that do not make logical sense. This can be a problem if there is overreliance on the mathematical fitting without supplementary expert judgement alongside this.
- Avoiding the above requires some degree of expert intervention, which may lose some of the automation benefits.

5.7.1.3 Information criterion approach

The above approaches could be seen as two extremes, and it may be that a firm’s approach may be somewhere in between. For example, the firm may conduct a fundamental analysis, and then include all the terms it is sure are required, and then also additional terms where it wants to investigate the significance of the term. It could then perform the large number of fitting runs required and use a mathematical approach to reduce the form until it is optimal.

Firms who have more development time may be able to start with a theory driven approach, and then, through repeated validation of the model, gradually include additional terms. Conversely, a firm that needs to establish a validated model more quickly may feel that the additional processing time for fitting runs is worthwhile to enable a data-driven approach to fitting, as this could be achieved in one iteration.

Whatever approach is taken, firms will need to carefully document the judgements made. It is important that firms retain a line of sight between the forms of their proxy models and the financial risks that they are exposed to. Therefore, whatever approach is used, firms should be able to communicate to senior management the method they have adopted and why it suitably covers the risks the company faces.
5.7.2 Choice of fitting points

In the above section, we described high-level approaches to fitting the model. The approach taken will influence the choice of fitting points to some extent, because we can only fit certain forms of function with a minimum number of fitting points. Beyond this, to achieve the objective of an accurate proxy model, the firm will have a number of further considerations:

- The main constraint will be the capacity of a firm’s heavy models and related processes to compute and analyse stress scenarios. There may be insufficient fitting data to select a proxy function using data alone. This may lead to firms relying more on theory. Firms will need to allocate the fitting runs optimally, to avoid wastage. In order to run the fitting points and fit the model in time for its deployment, there is likely to be a ‘budget’ for the number of fitting runs.

- Firm should also decide on the relative number of univariate and multivariate points. Multivariate points could be randomly generated risk sets. Multivariate points can provide more information on risk interactions, but are generally harder to validate not least because they are harder (or impossible) to visualise. Other firms may choose to run large amounts of univariate points, and run targeted bivariate (or triple, quadruple, etc.) risk combinations based on expert judgement of the likely interactions. Validation of these scenarios is generally easier, but the firm runs the risk of missing out on important interactions.

- In principle, under a fully data-driven approach it will be possible to extract more information from a given ‘budget’ of fitting runs when these are randomly generated or based on low-discrepancy sequences. Firms may have access to a large number of univariate stresses performed for other purposes and these can be used to supplement the ‘budget’.

- Where univariate stresses are used, only a handful of fitting runs will provide any information as to the relationship between the risk metrics and a particular risk driver. In contrast, where randomly generated multivariate stresses are used, all fitting runs will provide some information for all risk drivers.

- Fitting points can generally be randomised or manually selected, for example to cover certain percentiles. Randomised points reduce the reliance on expert judgement, but run the risk of missing out on key fitting areas through chance. Manual selection can force an adequate range of fitting. Randomised fitting approaches could be more appropriate if the ‘fitting budget’ is high, as this will reduce the potential for missing areas of fit.

- The range of fitting points will partly depend on the use of the model. Most firms will be looking to create a full PDF, and so will want to achieve good fit to both sides of the distribution and across the full range. For more bespoke uses of the model, a more targeted fitting approach may be valid, for example to manage catastrophe risk there may be additional focus on the very extreme end of losses.

- Firms may also choose to vary the number and location of fitting points between different areas of the business e.g. entities or products, such that the intensity of the fitting work is proportional to the materiality. This would reflect the material risks for each sub-group, and may for example mean that firms only aim to fit linear curves to some areas, or ignore joint terms.

5.7.3 Fitting method

5.7.3.1 What level to fit the proxy functions to

Section 5.5 discussed how firms would determine the structure of the proxy model, including the organisational hierarchy. This would generally see Groups or Legal Entities at the top of the structure and some more granular categories at the bottom, for example products or asset portfolios. Once this structure is set, firms can also decide on what level they will fit the proxy functions.
It will be necessary for the most granular lower level functions to be fit in order to produce results for that level – otherwise the information criterion will not include any terms relating to that level in the proxy function. At higher levels, firms have two general choices:

a) Set the higher level loss functions to be the aggregate of lower level functions. This effectively means all the fitting is performed at granular level, and that any higher level aggregations are just the sum of these. Firms should be aware of how lowest level fitting errors accumulate up. In some cases, errors at the lowest level may offset to give a better fit higher up, but in some cases small errors at the lowest level may add up to something material at higher levels. One way to avoid this would be to set tighter limits at the lower levels, however this may just make successful fitting implausible.

b) Fit the model at different levels for different purposes, for example a Group economic capital calculation could be based on a different calibration to Legal Entity results. This way a good fit can be ensured at each reported level. The key disadvantage is that the results at different levels will be inconsistent. Further additional fitting work needs to be performed, and the simulation software used may not permit this approach. If the fitting is performed alone at a higher level, this would potentially compromise the fit lower down, this rendering those results invalid.

Therefore, most firms are likely to adopt (a), in order to avoid having different models for different levels of calculation. However, if using this approach, firms will need to ensure that the model is validated for all the levels at which it is used.

5.7.3.2 Fitting tools

- Regression methods

The most common method in practice is Ordinary Least Squares (OLS), which, for a given formulaic structures chooses the coefficients that minimise the Sum of Squared Errors. As a variant, firms may wish to use a Weighted Least Squares approach, which applies a weighting function to the regression to put more/less weight on different parts of the fit.

Hursey et al (2014) provide formulaic detail on these methods, and this is not repeated here.

- Stepwise regression

A mathematical technique would mean adopting a stepwise approach to fitting the model. The aim here is to step through different possible structures of the proxy function – in terms of which terms and cross-terms are included – and assess the fit at each level according some pre-defined statistical criterion. The steps can either start at a very basic model and build up additional terms, or vice-versa. The stepping through can be automated computationally and so implicitly allows the firm to consider a very large range of possible model calibrations and automatically choose the most appropriate based on that statistical criteria.

Different statistical criteria exist for use in a stepwise regression. The most commonly used in actuarial models are the Akaike and Bayes Information Criteria (AIC and BIC). These are both statistics that aim to assess the trade-off between the complexity of the model and its goodness of fit. More details can be found in appendix A.6.2.

Whatever method is used, graphical representation of the fit can also be instructive. For example this may allow us to assess the relative quality of the fit over the distribution. This may be important if we are willing to sacrifice fit in some areas to benefit others. The QQ plot is an example of this, whereby the fitted and actual values are plotted on each axis are shown again the optimal y=x line. Clearly graphical approaches cannot be
applied to any automated fitting, and as such are generally used to supplement the sorts of statistics listed above.

5.8 Validating Proxy Models

5.8.1 Introduction

In section 5.6.1, we set out the objectives for the fitting of the proxy model. They key objective is that the model is accurate in replicating the results from the heavy models. This section considers how insurers can test the proposed model and communicate the results of the testing to stakeholders to help justify the choices made.

5.8.2 Validation scope

The proxy model is usually part of a firm’s wider economic capital model, which may be used for regulatory purposes (e.g. Solvency II Internal Model) or for internal reporting (e.g. Economic Capital model). Multiple validation tools exist to validate these models, including back-testing, reverse stress testing and statistical testing. It is important to be clear what exactly is being tested and how the chosen validation tool design achieves the test objective.

Figure 5-3 shows the typical components of a capital model used in producing a Probability Distribution Forecast, and where the proxy model validation fits in. For example, under Solvency II, Internal Models are required to produce a suitably accurate PDF which quantifies the movement in Own Funds over a full probability distribution. The focus of this section is on statistical testing to validate the proxy model component and if relevant, how it is “rolled forward” prior to generating the PDF i.e. where the proxy model is initially calibrated prior to the valuation date, and then rolled forward to that date. The approach to roll-forward is discussed in section 5.8.11.

Hence the scope of this section is validating the components at points labelled (1) and (2) in Figure 5-3.

<table>
<thead>
<tr>
<th>Validation Point</th>
<th>What is being testing</th>
<th>Tool Ref</th>
<th>Tool</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1. How effective is the proxy model at replicating the results that would be achieved with the heavy model?
   - V1.1 Goodness of Fit tests
   - V1.2 Visual inspection of fitted curves
   - V1.3 Out of sample scenarios

2. How accurate is the roll forward methodology used?
   - V2.1 Out of sample scenarios
   - V2.2 Full back-test ‘out of cycle’

### 5.8.3 In-sample testing (goodness of Fit and visual inspection)

An element of validation should occur during the fitting process. After regression is used to fit the curves, firms should output a number of fitting statistics to inform how well they have been able to fit the observed points. This is likely to include:

- R-squared or R – a measure of overall correlation and explanatory power
- Mean-Squared Error (MSE), or Sum of Squared Error (SSE). Both measures of the absolute fitting error.
- Maximum (absolute) error
- Number of points outside a desired range – either an absolute or percentage amount.

These kinds of statistics are used commonly part of out-of-sample testing, so are discussed in the next section. Firms are also likely to use visual inspection to assess the fit. Statistical testing of in-sample points has limits on its credibility, because it does not cover the risk of un-modelled points being incorrect. If the user has sufficient expertise, graphical inspection can help to identify areas where the fit is inappropriate. For example, the fitted curve may show turning points outside the fitting range that we may not expect.

### 5.8.4 Out-of-sample testing

Out-of-sample testing is an example of an effective statistical process for validating the proxy model. Scenarios that are not used in fitting are evaluated using the heavy model and compared with the results using the proxy model. The same sorts of fitting statistics can be used as for in-sample testing. Graphical inspection of the fitted curves themselves may have limited use in out-of-sample testing because we are generally concerned with looking at multivariate points.

Out of sample testing tests for the impact of the following potential errors in proxy models:

<table>
<thead>
<tr>
<th>Error type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fitting error</td>
<td>Fitted functions do not fit well at certain fitting points (though this is part of in-sample testing).</td>
</tr>
<tr>
<td>Interpolation error</td>
<td>Shape between fitting points is inaccurate so that the fitted function does not predict the heavy model loss well in scenarios that are not close to a fitted point.</td>
</tr>
<tr>
<td>Extrapolation error</td>
<td>Behaviour of the fitted function beyond the last fitting point (but within the range of simulations) is inaccurate, for example through unintended turning</td>
</tr>
</tbody>
</table>
points or higher order polynomials increasing excessively at higher levels of the relevant risks.

| Interaction error | Combinations of risks are not accurately modelled or are not modelled at all. |

Proxy Models inherently contain an element of approximation, and as discussed in Hursey et al (2014), errors in individual points do not themselves invalidate the model. What is more important is that there is no bias in the fitted functions, so that we are neither systematically under/overstating a result, or failing to adequately rank risks. Proxy Model errors when assessing the PDF can be put into two categories:

1) Proxy functions are biased so that the magnitude of losses is incorrect, though the ranking may be broadly correct, for example where extreme losses are overstated by loss functions but would still be extremes in the ranking if the loss functions were accurate;

2) Proxy functions are biased so that both magnitude and ranking of losses is wrong. This can arise for example due to turning points calculating profits when there should be losses (and vice versa) or where interactions are inaccurately modelled such that some combinations of risk have effects which are completely unexpected, for example where a fit is accurate for scenarios when two risks both increase, but the wrong sign when the risks move in different directions.

In the former case, the model’s ability to rank risk may be appropriate, but the financial outcomes are not. For example Economic capital may be over/understated. Validation statistics would should that proxy model is systematically over/understating the heavy model evenly across the full PDF.

In the latter case, the validation statistics would show errors that are not similar across the full PDF.

The reminder of this section focusses on the choices that need to be made when conducting out-of-sample testing, and how to communicate the results.

5.8.5 Choosing scenarios for out-of-sample testing

5.8.5.1 Number of Scenarios

The Working Party is not aware of any specific approach to deriving a theoretically robust number of scenarios to test, for example a formula based on statistical significance. Such approaches rely on the underlying assumption of randomness of errors which does not necessarily apply when replicating an underlying model.

Clearly the more scenarios tested the greater the confidence in the view formed from the tests on whether or not the proxy model is a good representation of the underlying Heavy models.

The following principles are relevant in determining a suitable sample size:

<table>
<thead>
<tr>
<th>Principle</th>
<th>Number of scenarios expected to increase:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Materiality</td>
<td>for a large fund than for a smaller fund with similar risk profile</td>
</tr>
<tr>
<td>Complexity</td>
<td>for a complex Group structure with multiple reporting entities, each with a different ‘critical scenario’ profile.</td>
</tr>
<tr>
<td></td>
<td>for a fund with diverse range of risks and/or complex interactions as will want to capture many variations.</td>
</tr>
<tr>
<td>Use</td>
<td>where the proxy model is used extensively such that a wide range of the</td>
</tr>
</tbody>
</table>
pdf is relevant than if mainly used at the level of the 99.5\textsuperscript{th} percentile

<table>
<thead>
<tr>
<th>Cost / Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>while the benefit justifies the cost as out of sample testing requires significant effort when there is sufficient time, for example ‘out of cycle’ vs ‘in cycle’</td>
</tr>
</tbody>
</table>

A recent survey\textsuperscript{14} of nine UK firms showed that the number of scenarios tested ranged from 20 to 1000 (median 50). Clearly the number of scenarios should be proportional to the complexity of the business, but we would expect that the median is required in most cases. Most firms would want to perform more than this to be able to use stronger statistical inference, and this may become possible as technology moves forwards.

We expect the industry will refine its view of the suitable volume of tests as firms gain further insights from the proposed testing and then may also become more focused in the testing. Until then the approach is likely to be driven by the maximum achievable number within reasonable constraints.

5.8.5.2 Which Scenarios?

It is important to test the full range of the scenarios across the PDF because different regions and intensities are relevant for different uses of loss functions. There is also a strong reference to this in the regulatory requirements mentioned earlier.

Because we are looking to test the PDF, which is a representation of the full distribution of all the risk factors to which a company is exposed, the emphasis should be on testing multivariate points covering all risks. This ensures that the points tested can be mapped to the overall percentiles (and in fact are likely to be derived from percentile outputs of the model).

Scenarios that consist of a subset of risks can be appropriate when we are looking at a specific section of the model (e.g. asset, product or entity), which may be the case if the model is being used for a specific purpose, or if there investigations that need to be performed after issues have been identified with the multivariate scenarios.

Notwithstanding the need to assess the full Probability Distribution Forecast, firms may choose to model a higher concentration of points at different percentiles, as ultimately different points will be more relevant for different uses of the model. The following table sets out an example target allocation of the number of scenario across percentiles for a reporting entity. For this purpose the percentile range has been split into 6 sections broadly aligned to usage. The scenarios are allocated across these percentiles in each fund to test various uses.

<table>
<thead>
<tr>
<th>Probability Distribution Region</th>
<th>Broad Percentile range</th>
<th>Specific Uses</th>
<th>Allocation of scenarios</th>
</tr>
</thead>
</table>
| SCR/Economic Capital region     | 99\textsuperscript{th}-99.9\textsuperscript{th} | i) SCR calculation  
                                  |                        | ii) Diversified risk levels  
                                  |                        | iii) Risk management prioritisation | 50% |
| Mid to High adverse            | 70\textsuperscript{th}-99\textsuperscript{th} | Risk Appetite / Stress and Scenario testing (e.g. cover 1in5 or 1in10 events) | 10%-20% |
| Low to Mid adverse             | 50\textsuperscript{th}-70\textsuperscript{th} | Short-term solvency estimates using “roll forward techniques”. Both adverse and beneficial movements are | 10%-20% |
| Low to Mid beneficial          | 30\textsuperscript{th}-50\textsuperscript{th} |

\textsuperscript{14} Reference: Deloitte survey 2014
### SCR/Economic Capital

The simulations around the 99.5\(^{th}\) percentile are most relevant for calculating the aggregate 1 in 200 capital requirement. The SCR or other Economic Capital measures are typically calculated as the 99.5\(^{th}\) (or similarly high percentile) ranked loss across a large number of simulations. SCR will be understated if more scenarios ranked at a lower percentile than 99.5\(^{th}\) are undervalued relative to scenarios ranked at a higher percentile or if there is a systemic error across many scenarios. The extent of the movement in SCR will depend on how far undervalued the incorrect simulations are and how close they are to the 99.5\(^{th}\) ranked loss.

To test ranking errors, simulations near the 99.5\(^{th}\) percentile are important. The further away from this region the greater the error needs to be to influence the ranking, although large errors are still possible where turning points exist. Systemic errors can be detected from anywhere in the range.

### Diversified Risks (or allocated capital)

Diversified losses by risk are referenced for a number of purposes, including materiality definitions and risk appetite and prioritisation for risk management. A typical method of calculation is to take the average level of risk across simulations ranked from say 1,000 simulations around the 99.5\(^{th}\) percentile, possibly weighted to give more weight to simulations near the 99.5\(^{th}\). The range of 1,000 simulations is often known as “critical region” or “smoothing window”.

The accuracy of losses across the smoothing window is important to the accuracy of the diversified losses and scenarios.

### Short-term solvency estimation

Proxy models are commonly used to estimate changes in the economic balance sheet position between full calibrations. Typically these are smaller movements that are actually experienced, both positive and negative. Testing scenarios at other percentiles will provide useful information on the accuracy of loss functions in scenarios that might be encountered for this purpose. There is no specific percentile that needs to be focused on for this purpose, but a range of levels positive and negative should be tested.

### Risk appetite limits

Some insurers use proxy models to periodically set the level of capital needed to cover SCR after a 1 in X level event, typically a 1 in 5 (80\(^{th}\) percentile), a 1 in 10 (90\(^{th}\) percentile) and a 1 in 25 (96\(^{th}\) percentile). Proxy models are used to estimate the impact of these events on Own funds and on SCR by assessing a 1 in 200 loss after the 1 in X event.

Therefore the adverse range below the 99.5\(^{th}\) smoothing window needs to be tested for the Own Funds impact. Scenarios above 99.5\(^{th}\) may be relevant to validate the capability of loss functions to calculate the SCR after a 1 in X event.

### Stress and scenario testing (SST)

SST can use proxy models to assess own funds and SCR impacts after a defined scenario – which may be a 1 in X scenario generated or a real-life defined scenario such as a pandemic. The testing in regions described for risk appetite setting is also suitable for validating SST.

#### 5.8.5.3 Scenario Selection Method

Scenarios may be either selected to target specific combinations of risks or selected non-systematically, such as at random or from pre-defined percentiles of the aggregate distribution.

Examples of the specific selection approach include:

- Targeting specific key interactions in each fund
• Targeting outliers e.g. large credit scenarios
• Targeting known limitations / poor fits

These approaches have the following challenges/disadvantages:

• They introduce potential bias towards poor fits.
• There are many potential permutations of interactions and probability levels that theoretically might be relevant. For example a test pack for an annuity fund could include various pairs or triples of interest rates level, interest rate slope, longevity and credit and then with each at different probability levels and directions. It is difficult to judge the most appropriate and efficient subset of the permutations for testing. Over time, repeating the validation on new sets of simulations should build up a body of evidence that will test a wide range of these permutations and therefore continue to enhance understanding and help further refine the loss functions.
• It is manually intensive to find simulations that exhibit the required features and it can be the case that very few simulations in the relevant region exhibit material occurrence of particular interactions targeted.

5.8.6 Acceptance criteria

In order to assess whether the proxy model is fit for purpose, there is a need to define what is an acceptable deviation from the heavy model results i.e. a validation “pass” or “fail”.

The selection of tolerances is one of the most important steps in the fitting and validation of proxy models. Tolerances determine the level of effort required in terms of the number of valuation model runs and the required complexity of the proxy model. In general, the lower the tolerances the higher will be the number of scenarios required in fitting and validation.

Setting arbitrarily tight tolerances may be spurious accuracy that will drive significant additional cost without altering management decisions. Setting too high tolerances may lead to incorrect decisions.

5.8.7 Validation measures

Firms should consider a range of statistics when analysing their out-of-sample results. This should be a combination of descriptive and analytical statistics, and can include:

Measures of fit for individual samples:
• Monetary error (versus heavy model) in relevant currency
• Percentage error e.g. as percentage of true heavy model result
• Absolute error

Descriptive measures for the full sample set:
• Number of samples outside a specified range (defined ad monetary value)
• Number of samples outside a specified range (defined as percentage)
• Mean Squared Error
• Mean Absolute Error

Analytical statistics:
• R-squared\textsuperscript{15} (coefficient of determination). Measures how much of the variation in the heavy model result is explained by the heavy model. A value close to 1 (say > 0.99) is expected if the proxy model is a good predictor of the heavy model.
• R (Pearson Product-Moment Correlation Coefficient). This measures the linear dependence between values from the fitted function and the in-sample modelled values.\textsuperscript{16}
• Coefficient of Skewness: This is a measure of the extent to which the errors are biased in either direction (over/understatement of heavy model result). This can tell us at an overall level whether the proxy model tends to under or overstate results.

For the majority of these statistics, we can view the results for the entire scenario, or for specific regions of interest. This is particularly pertinent when looking at the skew of the errors. The skewness statistics could be low for the distribution as whole, but may contain clustering of bias at certain points of the curve. This is where graphical inspection of the out of sample points can be instructive, since patterns in the residuals can be plotted and assessed accordingly.

5.8.8 Setting Tolerances

One way of setting tolerances is a top-down approach starting with management’s acceptability for the level of inaccuracy for each use of the proxy model and cascade down to individual scenario level as follows:

1) Agree the level of accuracy required so that the model ranks risk sufficiently accurately and help senior management to make business decisions. What level of error in the key output e.g. SCR, Risk Appetite levels would drive management to alter a decision. This is aligned to the definition of materiality in Article 222 of the Draft Solvency II Delegated Acts, dated July 2014

“For the purposes of this Chapter, a change or error in the outputs of the internal model, including the Solvency Capital Requirement, or in the data used in the internal model shall be considered material where it could influence the decision-making or the judgement of the users of that information, including the supervisory authorities.”

2) At the same time the tolerances should be set in such a manner that the effort required in achieving them is a) proportionate, b) not unduly onerous to implement, c) makes allowance for accumulation of errors. This may lead to a simple high-level % error.

3) Based on this high level tolerance, design the tolerances at lower levels e.g., legal entity level or sub-fund level or at individual risk level.

Tolerances may differ in different regions of the Probability Distribution Forecast reflecting that the SCR / diversified risk levels are likely to drive more risk management actions than say errors in very extreme scenarios. However this may overcomplicate the validation operationally and require more scenarios in each region to get a representative view.

Some examples of tolerances that could be used for the statistics derived in section 5.8.7 are shown next.

5.8.9 How should the errors be investigated?

Individual errors can flag types of risk combination that are particularly inaccurate e.g. due to a missing or inaccurately modelled interaction. It may be that such combinations are rare and therefore the fail is not particularly important or it may be that the error occurs in many simulations and therefore causes errors across the Probability Distribution Forecast.

\textsuperscript{15} R^2 = 1- \text{Sum of squared Errors}/\text{Total Sum of Squares}

\textsuperscript{16} Note – in this case we can use either R or R^2 and not both, since the relationship we are trying to test is effectively an X = Y relationship, and there is no intercept in the regression.
The cause of the failing individual scenario should be understood or at least the unique features of the scenario compared with the passing scenario should be identified and the simulation set examined to establish how frequent that type of scenario is to determine if further remedial action is justified due to pointing to a more systemic issue.

5.8.10 Communicating validation results

The validation results can be communicated in the form of a QQ plot showing the probability level of the chosen scenarios and how they relate to uses of the Probability Distribution Forecast (see Figure 5-4). It is helpful to use a consistent format of presentation as used when setting the scenarios to test (see Section 5.8.3).

![Figure 5-4 Quantile-quantile (QQ) plot of proxy model errors against heavy model results](image)

This plot can be used to communicate the suitability of the proxy models to the users of the model for each of the use types highlighting the percentiles and situations where the fit of the model is good and where it is poor/uncertain and the impact that can have on the key output for each model use. This will help management understand the limitations of the model and therefore assist the user to adjust use accordingly rather than over rely on it where it is less accurate.

The graphical plots can also assist in understanding the distribution of residuals across the range of percentiles. This can bring out any issues that are present in certain areas of the curve, but which are lost when statistics are calculated for the full sample set (such as, for example the outliers shown on Figure 5-4).

5.8.11 Roll Forward

5.8.11.1 Drivers

Roll forward is a fairly common practice used in producing the Probability Distribution Forecast at the valuation date. It involves fitting the proxy model at an earlier date (for example end of Q3) and then adjusting the fitted
model to the valuation date (end Q4) using approximate scaling factors. This approach enables SCR to be calculated, used and reported earlier than would otherwise be the case. This is an important consideration for management reporting economic capital for multiple reasons including:

- To inform timely management interventions to manage risks
- To meet expectations of the market for external reporting in the case of shareholder owned companies
- To meet regulatory reporting requirements.

The desired timescales may leave insufficient time to fully calibrate and validate proxy models using “hard-close” asset and liability data at the reporting valuation date. Instead actuaries have developed techniques that are designed to provide sufficiently accurate results with less dependency on “hard-close” data by “rolling forward” from previously calibrated models.

Typically these techniques involve:

- Calibration “off cycle” in advance of the reporting period
- Roll forward methodology to approximate the impact as at the valuation date of experience in the period since the calibration
- A trigger framework to flag a need to intervene should the approximations have exceeded tolerance for inaccuracy.

5.8.11.2 Roll forward methods

Roll forward methods include

- Scaling some proxy functions according to a readily available “carrier” for the movement in risk, for example policy counts, sum assured or annuity payments in force – essentially some approximate measure that does not involve a heavy model calculation (such as best-estimate liability). Such risk carriers may capture changes in volume or changes in sensitivity or both.
- Partial recalibration, for example to focus recalibration on lines of business that are most sensitive to risk factors, or recalibrate the most material risk factors only.
- Leaving risks unchanged

The choice of roll forward method for a particular risk depends on:

- The materiality of the risk
- How stable is the risk
- What circumstances might cause the magnitude of the risk to change materially

5.8.11.3 Validation Considerations

Roll forward will potentially introduce additional approximation error over the proxy model calibration errors discussed earlier. The tolerance, if based on the top-down approach, would logically require both sources of error to aggregate to within the tolerance rather than testing the two components separately.

Firms face choices about which point(s) to perform validation:

- Perform validation at the calibration date and then perform an analysis of the movement over the roll forward period. The latter would be considered ‘light’ validation, and could for example include stepping through each change to the proxy functions and risk factors to understand the cause of change and assess its reasonableness. Because this validation is light, it is best suited to more benign environments, for example where the exposure to risk is stale, there is a fairly stable mix of business, or when investment markets have been stable. Alternatively, if the roll forward period is short, then this level of validation may be appropriate.
Perform validation at the valuation date only. This will most directly validate the model at the point at which it is being used. However, depending on the timescales for reporting, this may be a significant challenge.

Perform validation at both the calibration date and valuation date. For example, firms could test intensively at the calibration date, then perform a smaller amount of testing at the valuation date before reporting deadlines.

The relative merits of each approach are discussed below.

Out-of-sample testing at date of calibration:

- A cleaner controlled validation as do not have roll forward approximation affecting the validation results,
- The validation can easily be completed “in cycle” before results are reported
- However this only the tests proxy model at the point of calibration. Roll forward would need to be validated separately; at a minimum and analysis of movement over the period can be used to back up the results from the proxy model. However in non-benign conditions, or where the roll-forward period is long, this analysis may be difficult to reconcile. Therefore, it may be necessary to produce some amount of out of sample testing at the valuation date.

Out-of-sample testing at date of reporting:

- This most directly validates the specific result reported, and it tests both proxy model calibration and roll forward combined
- It will be difficult to complete “in cycle” as we cannot start until roll forward results available, and reporting deadlines may be short.
- It can be harder to determine root cause of failing scenarios without significant additional work to isolate roll forward errors from proxy model errors.

Current practice varies across the industry. Most firms perform a substantial amount of testing at the calibration date. Many firms perform a smaller amount of testing at the valuation date. The key consideration is likely to be the length of the roll forward period. Firms who roll forward for longer periods, say 3 months or greater, are more likely to conduct further testing at the valuation date\textsuperscript{17}.

An alternative mitigant, for those firms who do not wish to conduct further significant testing at the valuation date, would be to investigate and set trigger points. These trigger points would be designed to set limits at which the firm would start to doubt calibration of the proxy functions. These trigger points could relate to a number of things, for example:

- Large market changes – market movements that are severe enough to cast some doubt as to the appropriateness of the proxy functions, particularly for assessing a 1 in 200 level of confidence after the market movement, which may be very deep in the tails of the original calibrated proxy function.
- Business experience – large levels of lapses or new business, in particular if these are concentrated on certain products such that scaling of proxy functions may not represent the true movement in exposure to risk.
- Other business activities, such as changes to investment policy, new or altered reinsurance treaties or hedging strategies. The existing proxy functions may not be easily adaptable for these types of event.

There will usually be a degree of subjectivity involved in setting any trigger points, so the firm should be realistic is answering the fundamental question of what would invalidate the calibrated proxy model.

5.9 What to do in the event of poor fit

The goodness of fit of the proxy models may be limited in a number of different ways:

\textsuperscript{17} Source: Deloitte Survey 2014 ‘Market Survey on Approach to Out-of-Sample tests’
- Particular products/risks do not fit well. This may be acceptable if they are not material – in which case maybe shouldn't be separated – or there are offsetting fitting errors between them. We should always however have mind to the use of the model. For example poor fit for individual products may be satisfactory at entity level for an SCR calculation, but for product level reporting e.g. assessing economic value-added, the model may fail in this respect. Also firms should be wary of improving fit in some areas but not others, as this can distort the quality of fit higher up the hierarchy.

- Poor fit at legal entity level could lead to misstated financial results, so needs to be rectified. It is also likely to imply poor fit at other levels of the aggregation e.g. product level.

- Fit may be of different quality along the Probability Distribution Forecast. We may be happy that the killer-scenario region fits well, however Solvency II requires a full and valid whole PDF and a correct tool for ranking risk. In any case, we may also use the model for uses that require a good fit at other points, for example more central percentiles if the model is used for roll-forward.

- The model may contain over-fitting to particular aspect. For example, it may include lots of cross terms and be fit using a high volume of multi-dimensional points, however this may compromise fit for univariate risks at important points. This may cause problems when monitoring individual risk exposures.

All of the above means that we need to define failure of the model well. Indeed there could be many forms of failure, clearly linking the nature of the failure to the use of that model. As a result, a proxy model may be accepted for certain uses but not others.

If the acceptance criteria are not met, the firm will want to try and remedy this, for example increasing the number of fitting points or manual intervening in any mathematical fitting exercise. However, this may take time to remedy and with accelerated reporting cycles it is highly likely that this will not be possible within the timescales of results submission. One mitigant of this would be to conduct curve fitting exercises 'off-cycle', as described in the roll-forward section, so that such issues can be ironed out in advance of the reporting cycle.

It may be the case that the model fails for a particular use, for example the fit may be inadequate around the centre of this distribution, but well fit in the tails – perhaps because fitting has concentrated on that area. In such a case, the reasonable decision would be to accept the model for use in tail risk calculations, but not more moderate stresses, such as one may apply when performing roll forward calculations. In such a case, a firm would have some time to recalibrate to achieve a better fit in the middle of the distribution.

Note that the reverse would likely present more problems i.e. inadequate tail calculations are likely to cause a problem in calculating an SCR in required timeframes. In such a case, there are a range of options that could be considered to deliver a result in the time required:

- Use a previous recalibration – it may be better to use a previous accepted calibration with roll-forward from that calibration date. However, one should be cautious with this as there may have been fundamental changes that have caused the failure for the latest calibration. This would imply that the previous calibration may not be a suitable reflection of the risk exposures. It may alternatively imply that there has been a mistake or otherwise poor choice in the latest fitting exercise, and it would be preferable to revert to the previous calibration. Note that it should be possible to only re-use the failing proxy functions, rather than the full set. Using historical calibrations moves us further from the 'live' position, so adequate out of sample testing should be performed on the 'rolled forward' loss functions (to the extent this is possible in the time provided).

- An alternative would be to move to a variance-covariance approach. By its nature this is more simplified, but would suffer from the issues discussed in section 3. In particular, it is avoiding a fitting problem by just making unrealistic assumptions e.g. with-respect to linear exposures. This could be remedied somewhat through non-linearity adjustment, although again the methods for deriving the biting scenario are imperfect.
In all, it is never clear that a variance-covariance approach would provide a more appropriate result than a copula approach with invalid loss functions – though clearly this depends on the size of the fitting errors observed.

- A problem for Economic Capital/SCR could be addressed through a killer-scenario run through the heavy models for part or all of the relevant business. However, it is difficult to justify using an observed killer-scenario from rejected stochastic model, unless that part of the model that fails is unlikely to material change the composition of any (smoothed) killer scenario. It may be possible to look at historical killer-scenarios from validated models, but only if we feel there should not have been a material change in the risk profile.

- It may be necessary to add (normally prudent) adjustments to the results calculated, to reflect the invalid fit. But how could these be calculated?

  o A firm could adjust the out of sample results directly e.g. what was the worst (prudent) fitting point for the metric of interest e.g. worst within smoothing window.
  o Alternatively, a firm could adjust the previous calibration and results from that model, plus use any Analysis of Movement or roll forward information available i.e. what does the firm expect the result to be? Again may be supplemented by prudent adjustments.

- The out of sample error observations themselves can be used:
  o The out of sample scenarios could be added to the fitting set to improve the fit. However, the firm would need further out of sample points for validation, as using these additional points could potentially create new errors in other parts of the proxy model.
  o Out of sample errors can be used to estimate an impact on the metric of interest. In this case, the observed errors would be added as additional loss terms, which would then be used to estimate the additional loss in each copula simulation. This has the advantage of not requiring the firm to recalibrate its loss functions, and hence not distorting areas of the curve which may be well fit. However, the method required the model to interpolate between the out-of-sample points in order to estimate the loss for all other points. Hence it may be a more appropriate method in areas where we have a higher concentration of out-of-sample points, for example around the 99.5th percentile.

5.10 Addressing the communication challenges

Section 5.3 outlined some key communication challenges with respect to using proxy models. The above sections have described how proxy models can be designed, fit and validated. Building from that, the table below discusses how these communication challenges can be met:

<table>
<thead>
<tr>
<th>Communication Challenge</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Additional layer of approximation</td>
<td>Firms need to be clear on why a proxy model needs to be employed, including an analysis against potential alternative e.g. model simplification or heavy model point grouping. Model validation reports should be produced, which details the validation steps taken, and make solid conclusions about the appropriate uses of the model. It should be made clear if the model is not fit for certain purposes. Reporting to senior management should reference TASM and TASR, in making any limitations of the model clear.</td>
</tr>
<tr>
<td>Understanding the potential error around the true result</td>
<td>Senior management need to be aware that results from the proxy model are subject to some degree of error. It needs to be made clear what the potential error range is – this will be informed through communication of tolerance limits applied during the model fitting process, and the results of out-of-sample validation. In some cases, it may be pertinent to state a potential range of error around a financial result in a written report. Firm should use their judgement,</td>
</tr>
<tr>
<td>Understanding statistics and curve fitting</td>
<td>Senior management should receive specific training such that they have enough understanding of the proxy models, and their role in the financial results being calculated. Firms should ensure that relevant validation reports, or reports on results, contain accessible information in relation to proxy models. This is likely to result in avoiding complex statistics and mathematics as much as possible. Conclusions drawn from statistical analysis should be made very clear.</td>
</tr>
<tr>
<td>Range of model uses for proxy model</td>
<td>To the extent that the firm intends to use its proxy model for multiple purposes, the firm should ensure that the model is validated for that purpose, and that this validation is made clear to senior management in the report. Particular attention should be paid to the use of the proxy model in estimating balance sheet positions e.g. in Solvency Monitoring. As the true result will differ to some extent from the proxy result, the firm should investigate the differences and feed back into future calibrations.</td>
</tr>
</tbody>
</table>
6 References

<table>
<thead>
<tr>
<th>Ref</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGV</td>
<td>Delta-Gamma-Verfahren als StandardRisikomodell für Lebensversicherer</td>
</tr>
<tr>
<td>H</td>
<td>Higham, Nick “Nearest Correlation Matrix”</td>
</tr>
<tr>
<td>Ref</td>
<td>Reference</td>
</tr>
<tr>
<td>-----</td>
<td>-----------</td>
</tr>
<tr>
<td>Code</td>
<td>Author(s)</td>
</tr>
<tr>
<td>------</td>
<td>-----------</td>
</tr>
<tr>
<td>V1</td>
<td>Venter, G.</td>
</tr>
<tr>
<td>V2</td>
<td>Venter, G.</td>
</tr>
<tr>
<td>W</td>
<td>D. Wilkie</td>
</tr>
</tbody>
</table>
A. Appendix – Statistical Analysis

A.1. Definition: Pseudo-observations

Before launching into our example of the application of statistical techniques, it is useful to establish some notation.

Given a set of $N$ observations of a 1-dimensional random variable $\{X_1, ..., X_N\}$, the set of pseudo-observations $\{Y_1, ..., Y_N\}$ is defined by setting $Y_i = \frac{1}{N+1} \text{rank}(X_i)$ for $i = 1, ..., N$.

In other words, the pseudo-observations are the ranks of the random variables, scaled so that they take values in the interval $(0,1)$. It is usual to divide by $(N+1)$ rather than $N$ to obtain values which lie strictly between 0 and 1. This avoids singularities when using pseudo-observations in certain applications (e.g. Maximum Likelihood Estimation).

Where a random variable $X$ takes vector values, the pseudo-observations are obtained by applying the scaling above separately for each coordinate.

A.2. Data

Producing scatter charts of the data is a useful initial step. Charts can assist in:

- Identifying the presence of a relationship;
- The strength of the relationship (e.g. sign and magnitude of any correlation) and whether there are any patterns in that relationship for which an allowance may be appropriate (e.g. symmetry, any clustering of extreme values which may indicate the presence of tail dependence);
- Identification of any outliers

Charts of both the raw observations and pseudo-observations can be helpful. The latter filter out the marginal distributions and can be compared with scatter charts of standard copulas.
Figure 6-1 EQ/CR Raw Observations
Figure 6-2 EQ/CR Pseudo Observations

The charts show a negative correlation between EQ and CR. There also appears to be some clustering in the tails. The correlation between CR and PC1 also appears to be negative. It is less clear whether there is a correlation between EQ and PC1.

From the charts of the pseudo-observations, an elliptic copula for each of the pairs does not appear unreasonable.

A.3. Tests for independence

There are several statistical tests for independence available. Some of these are based on rank statistics (Spearman, Kendall and van der Waerden’s tests). Another tests whether the empirical copula is inconsistent with the independence copula using the Cramér von Mises statistic (see section C.4).

For all three risk factor pairs, unsurprisingly, all of these tests reject the null hypothesis that the pair is independent.

Further information on the tests is provided in Appendix C. The tables below show the results of the tests applied to our data.
Null hypothesis:

\[ H_0: \text{X and Y are independent} \]

### Spearman test

<table>
<thead>
<tr>
<th>Risk pair</th>
<th>Spearman’s rank correlation</th>
<th>95% CI Lower Bound</th>
<th>95% CI Upper Bound</th>
<th>p-value</th>
<th>Reject ( H_0 )?</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ/CR</td>
<td>-42.7%</td>
<td>-56.0%</td>
<td>-29.3%</td>
<td>3.89E-10</td>
<td>Y</td>
</tr>
<tr>
<td>CR/PC1</td>
<td>-28.1%</td>
<td>-41.4%</td>
<td>-14.7%</td>
<td>3.91E-05</td>
<td>Y</td>
</tr>
<tr>
<td>PC1/EQ</td>
<td>15.5%</td>
<td>2.1%</td>
<td>28.9%</td>
<td>0.02291</td>
<td>Y</td>
</tr>
</tbody>
</table>

### Kendall’s test

<table>
<thead>
<tr>
<th>Risk pair</th>
<th>Kendall's tau</th>
<th>95% CI Lower Bound</th>
<th>95% CI Upper Bound</th>
<th>p-value</th>
<th>Reject ( H_0 )?</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ/CR</td>
<td>-30.8%</td>
<td>-39.8%</td>
<td>-21.9%</td>
<td>1.57E-11</td>
<td>Y</td>
</tr>
<tr>
<td>CR/PC1</td>
<td>-19.2%</td>
<td>-28.1%</td>
<td>-10.2%</td>
<td>2.81E-05</td>
<td>Y</td>
</tr>
<tr>
<td>PC1/EQ</td>
<td>10.3%</td>
<td>1.3%</td>
<td>19.2%</td>
<td>0.024876</td>
<td>Y</td>
</tr>
</tbody>
</table>

### Van der Waerden’s test

<table>
<thead>
<tr>
<th>Risk pair</th>
<th>Standardised Statistic</th>
<th>95% CI Lower Bound</th>
<th>95% CI Upper Bound</th>
<th>p-value</th>
<th>Reject ( H_0 )?</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ/CR</td>
<td>-6.957408</td>
<td>-8.917372</td>
<td>-4.997444</td>
<td>3.47E-12</td>
<td>Y</td>
</tr>
<tr>
<td>CR/PC1</td>
<td>-4.491598</td>
<td>-6.451562</td>
<td>-2.531634</td>
<td>7.07E-06</td>
<td>Y</td>
</tr>
<tr>
<td>PC1/EQ</td>
<td>2.360983</td>
<td>0.4010193</td>
<td>4.320947</td>
<td>0.018227</td>
<td>Y</td>
</tr>
</tbody>
</table>

### Cramér von Mises test

<table>
<thead>
<tr>
<th>Risk pair</th>
<th>Test statistic</th>
<th>p-value</th>
<th>Reject ( H_0 )?</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ/CR</td>
<td>0.3933557</td>
<td>0.0005</td>
<td>Y</td>
</tr>
<tr>
<td>CR/PC1</td>
<td>0.1717645</td>
<td>0.0005</td>
<td>Y</td>
</tr>
<tr>
<td>PC1/EQ</td>
<td>0.0334665</td>
<td>0.0335</td>
<td>Y</td>
</tr>
</tbody>
</table>

#### A.4. Model fitting

We have assumed that, perhaps for the reasons described in section 5.6, the firm has opted to use a single tier of aggregation based on the Gaussian, Student’s t or IT copula.

There are various statistical techniques for estimation of the copula parameters that extend the Maximum Likelihood Method or Method of Moments which are familiar in fitting models for one dimensional random variables.
A.4.1. Maximum Likelihood approaches

There are two slightly different versions:

(a) **Inference from Margins (IFM) approach**

This approach assumes parametric models for each of the individual risk factors as well as the copula. The usual approach would be to express the likelihood of the joint distribution as a function in the parameters of the copula and the marginal distributions. This can result in a high dimensional space in which to seek a solution.

The IFM approach splits the optimisation process into two parts:

- The parameters of each of the marginal distributions are estimated first using maximum likelihood.
- The fitted parameters of the marginal distributions are then kept fixed and the likelihood is then maximised for the parameters of the copula.

The values of the copula parameters therefore depend on the models and parameters chosen for the individual risk factor distributions.

(b) **Maximum Pseudo-Likelihood (MPL)**

This method avoids making assumptions about the marginal distributions by using their pseudo-observations as inputs to the density function of the copula when forming the likelihood function. The resulting likelihood function then depends only on the parameters of the copula and, unlike the IFM approach, not on the assumed models and parameters of the marginal distributions.

For the purposes of our illustration, we have adopted the MPL technique.

A.4.2. Method of moments

There are several methods which are similar to the Method of Moments for one dimensional distributions. These approaches involve expressing statistics such as rank correlations as a formula of the parameters of the copula, calculating the value of those statistics for the observed data and inverting the formula to solve for those parameters.

For the d-dimensional Gaussian copula, one can calculate Spearman’s rho or Kendall’s tau for the data and invert to solve for the correlation parameter using the formulae below:

\[
\rho = 2 \sin \left( \frac{\pi \rho_S}{6} \right) \quad \text{Equation 1}
\]

\[
\rho = \sin \left( \frac{\pi \rho_T}{2} \right) \quad \text{Equation 2}
\]

Where \(\rho_S\) is Spearman’s rank correlation, \(\rho_T\) is Kendall’s tau statistic and the correlation matrix of the bivariate copula is \[
\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}
\]. See [ref. McNFE page 15]. Equation 2 holds more generally for an elliptic copula. Equation 1 does not extend beyond the Gaussian copula, although in practice, it does not appear to give significantly different results.

A similar approach based on Kendall’s tau can be adopted for 2-dimensional elliptic copulas such as the T copula. However, this only populates the correlation matrix. Another approach must be used to estimate the degrees of freedom parameter. The formula using Spearman’s rho does not extend to more general elliptic copulas. See [ref. McNFE Proposition 5.37 and Example 5.38].

The results do not generalise to higher dimensions. Nevertheless, both techniques may be used to estimate the correlation matrix (which may require further adjustment to make it positive semi-definite) before estimating the degrees of freedom parameters using MLE techniques.
One other possible approach to estimate the degrees of freedom parameter of a bivariate T copula is to estimate the coefficient of tail dependence $\lambda$ which is related to the correlations and the degrees of freedom parameter by the following equation:

$$\lambda = 2t_{\nu+1} \left[ -\sqrt{\frac{(\nu + 1)(1 - \rho)}{(1 + \rho)}} \right]$$

If the correlation has been chosen, the resulting equation can be inverted to obtain the degrees of freedom parameter. The coefficient of tail dependence is an asymptotic value and therefore not practical to derive from a limited volume of data. An alternative approach is to calibrate the parameters of the copula models to produce an appropriate value of the finite coefficient of tail dependence at a chosen percentile. This approach is described further in section 4.9.

We have fitted copulas on both a bivariate and a trivariate basis. The results are shown below.

**Maximum Pseudo-Likelihood**

### Bivariate – MPL

<table>
<thead>
<tr>
<th>Copula</th>
<th>EQ/CR</th>
<th>CR/PC1</th>
<th>PC1/EQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>$-48.8%$</td>
<td>$-31.8%$</td>
<td>16.8%</td>
</tr>
<tr>
<td>T</td>
<td>$-46.5%$</td>
<td>2.60</td>
<td>$-31.2%$</td>
</tr>
<tr>
<td>IT</td>
<td>$-40.3%$</td>
<td>2.50</td>
<td>1.68</td>
</tr>
</tbody>
</table>

**Trivariate – MPL**

<table>
<thead>
<tr>
<th>Copula</th>
<th>Correlation matrix</th>
<th>Nu1</th>
<th>Nu2</th>
<th>Nu3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>100.0%</td>
<td>$-48.8%$</td>
<td>100.0%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>17.1%</td>
<td>$-31.8%$</td>
<td>100.0%</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>100.0%</td>
<td>$-48.6%$</td>
<td>100.0%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15.0%</td>
<td>$-28.7%$</td>
<td>100.0%</td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>100.0%</td>
<td>$-46.5%$</td>
<td>100.0%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14.1%</td>
<td>$-29.3%$</td>
<td>100.0%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.52</td>
</tr>
</tbody>
</table>

We have used the inverse Kendall approach with the degrees of freedom parameter for the T-copula estimated using maximum likelihood.

### Bivariate

<table>
<thead>
<tr>
<th>inverseTau</th>
<th>EQ/CR</th>
<th>CR/PC1</th>
<th>PC1/EQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>$-46.6%$</td>
<td>-</td>
<td>$29.6%$</td>
</tr>
<tr>
<td>T</td>
<td>$-46.6%$</td>
<td>2.60</td>
<td>$29.6%$</td>
</tr>
</tbody>
</table>

### Trivariate

---

95
### Copula and Correlation Matrix

<table>
<thead>
<tr>
<th>Copula</th>
<th>Correlation Matrix</th>
<th>Nu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>100.0%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-46.6% 100.0%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16.0% -29.6% 100.0%</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>100.0%</td>
<td>4.47</td>
</tr>
<tr>
<td></td>
<td>-46.6% 100.0%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16.0% -29.6% 100.0%</td>
<td></td>
</tr>
</tbody>
</table>

- For EQ/CR, the MPL and MoM fit for the T-copula produce a very low degrees of freedom parameter. This supports the presence of tail dependence evident in the clustering in the tails of the charts in section 4.7.2.
- For CR/PC1 and PC1/EQ, the fitted degrees of freedom parameter is higher than for EQ/CR, although MPL and MoM produce less consistent results. This also supports the presence of tail dependence, though it would be appear to be weaker in extent than for EQ/CR.
- The IT copula fit shows a reasonable degree of symmetry in the degrees of freedom parameters for EQ/CR. The relationship is more skewed for CR/PC1 which may be an indication that a T-copula may not be the most appropriate. Fit. The relationship is very highly skewed for EQ/PC1 with one large degrees of freedom parameter and one very small. This may indicate that an elliptic copula will be a poorer fit for EQ/PC1.
- In the trivariate case, the fits for the T-copula and IT copula show that the fitted degrees of freedom parameter vary depending on the “level” at which the copula is fitted. It is not possible to set these assumptions independently for each pair of risk factors. The MPL fitting process arrives at a parameter which is in some sense an “average” of those calculated separately for pairs of risk factors.

### A.5. Goodness of fit tests

#### A.5.1. Introduction

There are several tests available to test the goodness of fit of copula models. Some of these are analogues of goodness of fit tests for 1 dimensional distributions and involve testing a 1 dimensional statistic derived from the data. Others involve testing the data directly against the proposed copula model.

The table below lists the tests we have used, the copula assumed under the null hypothesis and references to where more detail may be found.

<table>
<thead>
<tr>
<th>Test</th>
<th>Copula</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mardia</td>
<td>Gaussian</td>
<td>Appendix A.5.4 or [ref. McNFE Section 3.1.4]</td>
</tr>
<tr>
<td>Kolmogorov–Smirnov and Anderson-Darling</td>
<td>Gaussian and T</td>
<td>Malverge and Sornette [ref. MS] Kole, Koedijk and Verbeek [ref. KKV]</td>
</tr>
<tr>
<td>Cramér von Mises blanket tests</td>
<td>Arbitrary</td>
<td>Genest and Rémillard [ref. GR]</td>
</tr>
</tbody>
</table>

These tests are described in more detail in appendices A.5.4 through A.5.7.

#### A.5.2. Transformation of data

The tests are based on pseudo-observations derived from the data. In some applications, it may be appropriate to transform pseudo-observations to another scale. This may be necessary to apply a test which assumes a particular form of marginal distribution or in order to exploit specific properties of the hypothesised copula.

This section provides some definitions and outlines the geometric properties which some of the tests exploit.
A.5.2.1 Normalised pseudo-observations

These are obtained by applying the inverse cumulative distribution function of a standard Normal random variable to the pseudo-observations i.e. \( Z_i = \Phi^{-1}(Y_i) \).

A.5.2.2 Studentised pseudo-observations

For a specified degrees of freedom parameter \( \nu \), these are obtained by applying the inverse cumulative distribution function of a Student-T random variable with \( \nu \) degrees of freedom to the pseudo-observations i.e. \( Z_i = t_\nu^{-1}(Y_i) \).

A.5.3. Mahalanobis distance

Many of the copula statistical goodness of fit tests for elliptic copulas make use of the geometric properties of elliptic copulas. One of the parameters of an elliptic copula is a covariance matrix. This covariance matrix defines an inner product and a measure of distance between two observations \( X \) and \( Y \) as follows:

Inner product: \( \langle X, Y \rangle = X^T \Sigma^{-1} Y \)

Distance \( d(X, Y) = \langle X - Y, X - Y \rangle = (X - Y)^T \Sigma^{-1} (X - Y) \)

The inner product \( \langle X, Y \rangle \) is known as the “Mahalanobis inner product” between \( X \) and \( Y \) and the distance \( d(X, Y) \) as the Mahalanobis distance between \( X \) and \( Y \).

In particular, the squared Mahalanobis distance \( D \) of \( X \) from its mean \( \mu \) is defined by:

\[
D^2 = (X - \mu)^T \Sigma^{-1} (X - \mu)
\]

Equation 3

Where \( \mu \) is the mean of \( X \) and \( \Sigma \) is the variance/covariance matrix of \( X \).

It can be shown that if \( X \) follows a d-dimensional multivariate Normal distribution with mean \( \mu \) and positive-definite variance/covariance matrix \( \Sigma \), then \( D^2 \sim \chi^2_d \). (See [ref. McNFE Equation 3.14 or JW Result 4.7]).

A.5.4. Mardia’s test

A.5.4.1 Overview

Mardia’s test for multivariate Normality is a generalisation of the Jarque-Bera test for Normality in 1 dimension. Assume we have a set of \( N \) observations of d-dimensional vectors which are hypothesised to come from a d-dimensional multivariate Normal distribution. Mardia’s test is based on two test statistics \( b_d \) and \( k_d \) which are measures of the skew and kurtosis of the distribution. The values of these test statistics are based on observed values of:

(i) the squared Mahalanobis distance from the sample mean;
(ii) the Mahalanobis inner product between different observations in the sample relative to the sample mean.

Define:

\[
D_i^2 = \langle X_i - \bar{X} \rangle^T S^{-1} (X_i - \bar{X})
\]

\[
D_{ij} = \langle X_i - \bar{X} \rangle^T S^{-1} (X_j - \bar{X})
\]

where \( \bar{X} \) is the sample mean (a d-dimensional vector) and \( S \) is the sample variance/covariance (a \( d \times d \) matrix) defined by:

\[
\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i
\]

\[
S = \frac{1}{N} \sum_{i=1}^{N} (X_i - \bar{X})(X_i - \bar{X})^T
\]
The test statistics are then defined as follows:

\[
b_d = \frac{1}{n^2} \sum_{i=1}^{N} \left( \sum_{j=1}^{N} D_{ij}^3 \right)
\]

\[
k_d = \frac{1}{n} \sum_{i=1}^{N} D_i^4
\]

Under the null-hypothesis of multivariate Normality, the asymptotic distributions of the test statistics are as follows:

\[
\frac{1}{6} Nb_d \sim X_{d(d+1)(d+2)/6}^2
\]

\[
k_d - d(d + 2)
\]

\[
\sqrt{8d(d + 2)/n} \sim N(0, 1)
\]

A.5.4.2 Application in practice

Under the null hypothesis “H0: the copula underlying the distribution of a d-dimensional random variable \((X_1, ..., X_d)\) is Gaussian”, the distribution of \((Z_1, ..., Z_d) = (\Phi^{-1}(F_1(X_1)), ..., \Phi^{-1}(F_d(X_d)))\) is multivariate Normal. This provides a version of Mardia’s test to provide a goodness of fit test for a Gaussian copula as follows:

(i) Derive the pseudo-observations from the sample data – this provides a non-parametric way of filtering out the marginal distributions.

(ii) Normalise the pseudo-observations using the transform described in appendix A.5.2.1.

(iii) If the underlying copula is Gaussian, then the meta-distribution formed in step (ii) is Normal

(iv) Apply Mardia’s test to the Normalised pseudo-observations derived in step (ii).

Mardia’s test has been implemented in the R package “MVN” using the function “MardiaTest”.

A.5.5. Kolmogorov-Smirnov and Anderson-Darling tests

A.5.5.1 Introduction

It is possible to use the statistic defined by Equation 3 (or an appropriate variant) to convert observations from a d-dimensional random variable to observations from a 1-dimensional random variable. Using the Gaussian or T copula as the null hypothesis, this 1-dimensional random variable will follow a certain distribution derived by exploiting the geometric property described in appendix A.5.3. It is then possible to calculate Kolmogorov-Smirnov and Anderson-Darling test statistics for this 1-dimensional sample (of which there are several variants) to test whether the observations are consistent with the null hypothesis.

A test for the Gaussian copula was described by Malvergne and Sornette (see [ref. MS]) and extended by Kole, Koedijk and Verbeek to the T-copula (see [ref. KKV]). Each of these tests adopts a similar approach and is described in more detail below.

A.5.5.2 Test of Malvergne and Sornette (MS) for the Gaussian copula

This test is described in detail (see [ref. MS]).

The test is based on the fact that, under the null hypothesis that the distribution of \(X\) is a d-dimensional multivariate Normal then the square of the Mahalanobis distance from the mean (Equation 3) is distributed as a \(\chi^2_d\) random variable.
The test proceeds in four phases as described below:

**Phase 1: Generate the Normalised pseudo-observations of the sample**

(i) Generate the pseudo-observations \( \left( Y^i_1, ..., Y^i_d \right) = \frac{1}{N+1} \left( \text{rank}(X^i_1), ..., \text{rank}(X^i_d) \right) \) from the sample data. Here the ranks are taken for each component separately – i.e. the ranks of the \( X^i_1 \) are assigned, disregarding the values of the other components. Then the ranks of the \( X^i_2 \) are assigned, disregarding the other components, and so on. Under the null hypothesis, this should be a sample of \( N \) observations from a Gaussian copula.

(ii) Normalise the pseudo-observations \( \left( Y^i_1, ..., Y^i_d \right) \) using the transform described \( \left( Y^i_1, ..., Y^i_d \right) \rightarrow \left( \Phi^{-1}(Y^i_1), ..., \Phi^{-1}(Y^i_d) \right) = (W^i_1, ..., W^i_d) \) described in appendix A.5.2.1. Under the null hypothesis, this produces a sample of size \( N \) from a \( d \)-dimensional multivariate Normal distribution where the marginal random variables \( W^i_j \) are not necessarily independent.

**Phase 2: Evaluate the KS or AD test statistic for the sample**

(iii) Generate the statistic \( (D^i)^2 = (W^i)^TS^{-1}W^i \) for \( i = 1, ..., N \), where \( S \) is the sample covariance matrix. Then, according to the general result of appendix A.5.3, the statistic follows a \( \chi^2_d \) distribution.

(iv) At this stage, we have \( N \) observations which, under the null hypothesis, come from a \( \chi^2_d \) distribution. We can then apply one of the several versions of the Kolmogorov-Smirnov or Anderson-Darling tests. These tests measure the “distance” between the empirical cumulative distribution function of the observed values and the cumulative distribution function of the distribution under the null hypothesis. A description of the measures of distance is provided below.

**Measures of distance**

Suppose we have \( N \) observations \( (X^1_1, ..., X^N_1) \) from a random variable \( X \) which is a Test Statistic. The Test Statistic \( X \) follows a distribution with cumulative distribution function \( F_{TS} \). We wish to test the hypothesis that the distribution \( TS \) is the same as a reference or target distribution \( Z \). There are various measures of “how far away” the observed values are from the assumed reference distribution. Large deviations of these measures from 0 would be an indicator that the Test Statistic did not follow the target distribution.

<table>
<thead>
<tr>
<th>Test</th>
<th>Test Statistic</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kolmogorov- Smirnov</td>
<td>KS1</td>
<td>( \max_x</td>
</tr>
<tr>
<td>Average Kolmogorov- Smirnov</td>
<td>KS2</td>
<td>( \int</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>AD1</td>
<td>( \max_x \frac{</td>
</tr>
<tr>
<td>Average Anderson Darling</td>
<td>AD2</td>
<td>( \int \frac{</td>
</tr>
</tbody>
</table>

The Anderson-Darling (AD) statistics are weighted versions of the Kolmogorov- Smirnov (KS) statistics and place more weight on observations in the tails. The KS statistics put equal weight on each observation.

---

18 It is also possible to assess the fit visually using Q-Q plots of the empirical and assumed distribution.
To apply these statistics to finite samples, the following approach is adopted:

Let \((Y_1, ..., Y_N)\) be the pseudo-observations of the test-statistic.

Then the value of \(KS_1\) is given by \(KS_1 = \max_i |Y_i - F_Z(x_i)|\).
The value of \(KS_2\) is given by \(KS_2 = \frac{1}{N} \sum_{i=1}^{N} |Y_i - F_Z(x_i)|\).

Analogous formulae apply for the AD statistics.

**Phase 3: Bootstrap the distribution of KS or AD test statistic**

Another subtlety is that, the covariance matrix of the target distribution has been estimated from the sample data. The statistics \((D^i)^2\) are therefore not exactly chi-square distributed. In order to allow for this, the p-values of the statistics derived in step (iv) above are obtained using bootstrapping techniques as described next.

**Bootstrapping**

The traditional Anderson-Darling test, which is similar to \(AD2\) but with the integrand replaced with its square, provides an analytic formula for its test statistic (AD) assuming that the distribution of \(Z\) is known a-priori.

\[
AD = \sum_{i=1}^{n} \frac{1}{n} \cdot \frac{2i}{n} \left( \log(F_T(x_i)) + \log(1 - F_T(x_i)) \right) - n
\]

However, in our case, the distribution of the target random variable is not known a-priori: we have to estimate the variance/covariance matrix from the sample data. This means that any standard analytical formula or tables must be used with some caution and it is preferable to use an empirical distribution for the Test Statistic generated using bootstrapping techniques.

By stage (iv) of the process described in Phase 2 above, we will have calculated one of the KS or AD test statistics for the sample. This is a scalar quantity. As discussed above, the reference distribution for the test statistic (\(KS_1, KS_2, AD_1, AD_2\)) is generated using bootstrapping techniques to calculate a p-value for the value of the statistic obtained from the sample.

The general idea behind bootstrapping is to generate a distribution of the statistic in question by simulating a large number of times from the assumed distribution. This allows an empirical distribution for the test statistic to be generated from which an estimated p-value can be derived.

In this particular case, the sequence after step (iv) is as follows:

- (v) Fix the sample variance/covariance matrix \(S\). This will remain constant throughout.
- (vi) Simulate \(N\) observations from a \(d\)-dimensional multivariate Normal distribution with covariance matrix \(S\).
- (vii) Apply steps (i) to (iii) to the new sample in (vi).
  This involves generating a new set of normalised pseudo-observations, calculating their sample variance/covariance matrix (which will generally be different from \(S\)) and using this to calculate a new set of values \((D^i)^2\) for \(i = 1, ..., N\).
- (viii) Recalculate the KS or AD Test Statistic from the \((D^i)^2\) from Step (vii).
- (ix) Repeat steps (vi) to (viii) a large number of times. This builds up an empirical distribution for the Test Statistic.

**Phase 4: Calculate p-value**

- (x) A p-value for the observed value in step (iv) can then be derived from the empirical distribution by calculating the proportion of values from the bootstrapped distribution which exceed the observed sample value derived at Step (iv) of Phase 2.
Note that it is appropriate to apply a one-sided test as it is large values of the Test Statistic that would lead us to reject the null hypothesis.

We are unaware of any standard package in R which performs this test. We have therefore developed used bespoke R code provided by Makin and Stevenson.

A.5.6. Test of Kole, Koedijk and Verbeek (KKV) for T-copula

The KKV test for the d-dimensional T-copula generalises the MS test for the Gaussian copula and utilises the elliptic symmetry of the T-copula. This test is described in detail (see [ref. KKV]).

The MS-test was based on applying Kolmogorov-Smirnov or Anderson-Darling tests to the statistic:
\[ D^2 = Z^T \Sigma^{-1} Z \]
where \( Z \) is the transformed vector \( Z = (Z_1, ..., Z_d) = \left( \Phi^{-1}(F_1(X_1)), ..., \Phi^{-1}(F_d(X_d)) \right) \) and the \( F_i \) are the cumulative distribution functions of the marginals.

The KKV test replaces this statistic with
\[ D^2 = Z^T \Sigma^{-1} Z / d \]

where \( Z = (Z_1, ..., Z_d) = \left( \tau_v^{-1}(F_1(X_1)), ..., \tau_v^{-1}(F_d(X_d)) \right) \) where \( \tau_v \) is the cumulative distribution function of the Student-T distribution with \( v \) degrees of freedom and \( d \) is the number of risk of factors.

Under the null hypothesis that the copula is a T-copula with correlation matrix \( \Sigma \) and \( v \) degrees of freedom, then \( D^2 \) is distributed as an \( F_{d,v} \) distribution.

This can be seen by noting that, under the null hypothesis, \( Z \) follows a T-distribution and therefore can be expressed in terms of a multivariate Normal mixing distribution:
\[ Z = \frac{1}{\sqrt{W/v}} X \]
Where \( X \) is a d-dimensional multivariate Normal and \( W \) is a \( \chi^2 \) random variable. \( D^2 \) can then be re-expressed as:
\[ D^2 = \frac{X^T \Sigma^{-1} X / d}{W/v} \]
i.e. \( D^2 \) is the ratio of a \( \chi^2 \) random variable divided by its degree of freedom parameter over a \( \chi^2 \) random variable divided by its degree of freedom parameter and so has an \( F_{d,v} \) distribution.

As for the MS-test, we replace the \( F_i(X_i) \) by the corresponding pseudo-observations and estimate the correlation matrix \( \Sigma \) from the sample data.

The test for the null hypothesis “H0: the copula is a \( t_v \) copula” proceeds as follows:

**Phase 1: Generate the Studentised pseudo-observations of the sample**

(i) Generate the pseudo-observations \((Y_1', ..., Y_d') = \frac{1}{N+1} \left( \text{rank}(X_1'), ..., \text{rank}(X_d') \right)\) from the sample data and produce an estimate of the correlation matrix \( \Sigma \) (e.g. using Maximum Likelihood Estimation).

(ii) “Studentise” the pseudo-observations using the transformation \((Y_1', ..., Y_d') \xrightarrow{\text{Studentise}} \left( \tau_v^{-1}(Y_1'), ..., \tau_v^{-1}(Y_d') \right) = (W_1', ..., W_d')\).

**Phase 2: Evaluate the KS or AD test statistic for the sample**
(iii) Form the statistic \((D_i)^2 = (W_i)^T S^{-1} W_i / d\) for \(i = 1, \ldots, N\). Under the null hypothesis, the \((D_i)^2\) for \(i = 1, \ldots, N\) should be a sample from an \(F_{d,v}\) distribution.

(iv) Evaluate the required KS or AD statistic for the sample \((D_i)^2\) under the null hypothesis that \((D)^2\) follows an \(F_{d,v}\) distribution. As some of the parameters of the target distribution have been estimated from the data, it is necessary to apply bootstrapping techniques to determine the distribution of the KS or AD test statistic to determine a p-value. The bootstrapping proceeds as described below.

**Phase 3: Bootstrap the distribution of KS or AD test statistic**

(v) Fix the sample variance/covariance matrix \(S\). This will remain constant throughout.

(vi) Simulate \(N\) observations from a \(d\)-dimensional multivariate T distribution with covariance matrix \(S\) and \(v\) degrees of freedom.

(vii) Apply steps (i) to (iv) above to the new sample in (vi). This involves generating a new set of normalised pseudo-observations, calculating their sample variance/covariance matrix (which will generally be different from \(S\)) and using this to calculate a new set of values \((D_i)^2\) for \(i = 1, \ldots, N\).

(viii) Recalculate the Test Statistic for the \((D_i)^2\) from Step (viii).

(ix) Repeat steps (vi) to (viii) a large number of times. This builds up an empirical distribution for the Test Statistic under the assumption that the null hypothesis \(H_0\) is true.

**Phase 4: Calculate p-value**

(x) A p-value for the observed value in step (iv) can be derived from the bootstrapped distribution obtained in step (ix).

**A.5.7. Blanket tests of Genest & Rémillard**

The MS and KKV tests rely on exploiting the geometric properties of the assumed copula to derive an intermediate 1-dimensional random variable and then apply goodness of fit tests to that 1-dimensional variable. They are therefore constrained by the choice of assumed copula model.

A number of academic papers relating to so-called “blanket tests” have appeared in recent years, notably those of Genest, Rémillard and collaborators. These “blanket tests” are not as restricted in the form of hypothesised copula and test statistics are derived directly from the copula without an intermediate transformation.

The blanket tests are parametric tests i.e. they assume that the true underlying copula comes from a family \(C_\theta\) of copulas parameterised by a parameter vector \(\theta\). For example, a bivariate t-copula is parameterised by a correlation assumption \(\rho\) and degrees of freedom parameter \(\nu\) so \(\theta\) in this case would be a two dimensional vector \((\rho, \nu)\).

The tests measure the “distance” of the empirical copula from the maximum likelihood estimate copula in the chosen family. If the distance is in some sense too large, then the hypothesis that the data comes from that family of copulas is rejected. We have used the goodness of fit test based on the Cramér-von Mises statistic (see [ref. GRB]). This test is implemented in the gofCopula function in the Copula package of R for families of copulas supported by that package. It uses the standard Euclidean (or “root mean square”) measure of distance:

\[
S_n = n \int \left[ C_{\hat{\theta}_n}(u, v) - C_n(u, v) \right]^2 dC_n
\]

where \(n\) is the number of observations in the sample, \(C_n\) is the empirical copula and \(C_{\hat{\theta}_n}\) is the estimator. The statistic can be written in a simpler form by replacing the arguments in the integral with the rank statistics:

\[
S_n = \sum_{i=1}^{n} \left( C_{\hat{\theta}_n} \left( \frac{R_i}{n+1}, \frac{S_i}{n+1} \right) - C_n \left( \frac{R_i}{n+1}, \frac{S_i}{n+1} \right) \right)^2
\]
The statistic is a measure of the deviation of the fitted copula from the empirical data across the whole distribution. It gives equal weight to each point.

The approach followed by the test (in the bivariate case) is as follows:

1. The data is converted into ranks \((R_i, S_i)\) and a pseudo-sample \((\tilde{U}_i, \tilde{V}_i) = \left( \frac{R_i}{n+1}, \frac{S_i}{n+1} \right)\) is generated.
2. The empirical copula \(C_n\) is generated
   \[ C_n(u, v) = \frac{1}{n} \sum_{i=1}^{n} I \left( \tilde{U}_i \leq u, \tilde{V}_i \leq v \right) \text{ for } (u, v) \in [0,1]^2. \]
3. The estimator \(\hat{\theta}_n\) is derived from the sample data.
4. The value of the statistic \(S_n\) for the sample data is calculated.
5. The distribution of the sample statistic \(S_n\) is then determined using bootstrapping techniques in order to generate a p-value. The following procedure is adopted:
   (a) Assuming that the true underlying copula is \(C_{\theta_n}\), a large number \(N\) of random samples (each of size \(n\)) are generated from the copula. These take the form \((\tilde{U}_{i,k}, \tilde{V}_{i,k})\) where \(i=1,...,n\) and \(k=1,...,N\).
   (b) For each of the \(k=1,...,N\) simulated datasets, the following quantities are determined:
      - The corresponding ranks \((R_{i,k}^*, S_{i,k}^*)\) and pseudo-sample \((\tilde{U}_{i,k}^*, \tilde{V}_{i,k}^*) = \left( \frac{R_{i,k}^*}{n+1}, \frac{S_{i,k}^*}{n+1} \right)\)
      - The empirical copula \(C_{n,k}^*(u, v) = \frac{1}{n} \sum_{i=1}^{n} I \left( \tilde{U}_{i,k}^* \leq u, \tilde{V}_{i,k}^* \leq v \right)\)
      - The estimator \(\hat{\theta}_{n,k}^*\)
      - A revised value of Cramér-von Mises statistic
      \[ S_{n,k}^* = \sum_{i=1}^{n} \left[ C_{\hat{\theta}_{n,k}}^*(\tilde{U}_{i,k}^*, \tilde{V}_{i,k}^*) - C_{n,k}^*(\tilde{U}_{i,k}^*, \tilde{V}_{i,k}^*) \right]^2 \]
6. The \(N\) sample values of \(S_{n,k}^*\) define the empirical distribution of the statistic \(S_n\). They can be ranked and, for a given confidence level \(\alpha\) the \(\alpha\)th percentile value chosen. If the observed value actually observed \(S_n\) is greater than this percentile, then the hypothesis that the copula is in the family \(C_{\theta}\) is rejected.

The test generalises in an obvious way to higher dimensions.

**A.5.8. Tests applied**

The copula package in R does not currently contain the IT copula. We have therefore not been able to carry out any goodness of fit tests for the IT copula. To do this would require development of bespoke code and may be included in the scope of further work.

The copula package in R also requires that the degree of freedom parameter in the goodness of fit test for a T copula is fixed value (i.e. the only “free” parameters in a family of T copulas are the correlation parameters). We have therefore tested integer candidates for the degree of freedom parameter on either side of the MLE fit.

---

19 \(I\) denotes the indicator function which takes value 1 if the condition in its argument is satisfied and 0 otherwise.
A.5.9. Results

The tables below show the results of the various goodness of fit tests described above. Tests which result in rejection of the assumed model at a 95% confidence level are highlighted in red. The conclusions are as follows:

Bivariate models

- A Gaussian model is not rejected for any of the pairs with the exception of EQ/CR and CR/PC1 under the Cramér von Mises test.
- The T-copula is also rejected for EQ/CR and CR/PC1 by the Cramér von Mises test and by the AD2 test for CR/PC1.
- p-values for the KS/AD tests and CvM tests are generally higher for the T-copula. This does not necessarily lead us to reject the Gaussian copula but reflects the fact that adding an additional parameter allows the model to fit the data more closely.

Trivariate models

- The Gaussian model is not rejected under all the tests with the exception of the Cramér von Mises test
- The T-model is not rejected under all the tests with the exception of the Cramér von Mises test and the AD1 test

Overall

- The majority of the tests do not result in rejection of a Gaussian model nor of a T-model.
- The Cramér von Mises test tests the copula model directly against the empirical copula without the use of an intermediate transformation. It would therefore be expected to be a stronger test. Both the Gaussian copula and the T-copula are rejected by the CvM test on a trivariate basis. Based on the bivariate results, this would appear to be as a consequence of a poorer fit for EQ/CR.

Bivariate

Gaussian copula

<table>
<thead>
<tr>
<th></th>
<th>EQ/CR</th>
<th></th>
<th></th>
<th>CR/PC1</th>
<th></th>
<th>PC1/EQ</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>Test stat</td>
<td>p-value</td>
<td>Test stat</td>
<td>p-value</td>
<td>Test stat</td>
<td>p-value</td>
<td></td>
</tr>
<tr>
<td>Mardia</td>
<td>0.120571</td>
<td>0.362</td>
<td>0.092569</td>
<td>0.504</td>
<td>0.019616</td>
<td>0.951</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8.078615</td>
<td>0.443</td>
<td>7.675228</td>
<td>0.275</td>
<td>7.875028</td>
<td>0.409</td>
<td></td>
</tr>
<tr>
<td>KS1</td>
<td>0.079109</td>
<td>0.131</td>
<td>0.046849</td>
<td>0.662</td>
<td>0.057239</td>
<td>0.418</td>
<td></td>
</tr>
<tr>
<td>KS2</td>
<td>0.026224</td>
<td>0.276</td>
<td>0.015351</td>
<td>0.713</td>
<td>0.019440</td>
<td>0.481</td>
<td></td>
</tr>
<tr>
<td>AD1</td>
<td>0.174282</td>
<td>0.283</td>
<td>0.168285</td>
<td>0.306</td>
<td>0.168218</td>
<td>0.293</td>
<td></td>
</tr>
<tr>
<td>AD2</td>
<td>0.065056</td>
<td>0.284</td>
<td>0.040264</td>
<td>0.722</td>
<td>0.049497</td>
<td>0.506</td>
<td></td>
</tr>
<tr>
<td>CvM</td>
<td>0.093596</td>
<td>0.000</td>
<td>0.083426</td>
<td>0.000</td>
<td>0.020489</td>
<td>0.441</td>
<td></td>
</tr>
</tbody>
</table>

T-copula fitted using MPL

<table>
<thead>
<tr>
<th></th>
<th>EQ/CR</th>
<th>2.1 DOF</th>
<th>CR/PC1</th>
<th>9.4 DOF</th>
<th>PC1/EQ</th>
<th>6.1 DOF</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>Test stat</td>
<td>p-value</td>
<td>Test stat</td>
<td>p-value</td>
<td>Test stat</td>
<td>p-value</td>
</tr>
<tr>
<td></td>
<td>KS1</td>
<td>KS2</td>
<td>AD1</td>
<td>AD2</td>
<td>CvM</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.061038</td>
<td>0.020719</td>
<td>0.208262</td>
<td>0.050182</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.473</td>
<td>0.564</td>
<td>0.203</td>
<td>0.613</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.030509</td>
<td>0.008719</td>
<td>0.300928</td>
<td>0.025729</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.971</td>
<td>0.990</td>
<td>0.022</td>
<td>0.973</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.050996</td>
<td>0.014390</td>
<td>0.215212</td>
<td>0.039076</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.548</td>
<td>0.772</td>
<td>0.103</td>
<td>0.759</td>
<td>n/a</td>
<td></td>
</tr>
</tbody>
</table>

**T-copula (DOF parameter rounded down from MPL value)**

<table>
<thead>
<tr>
<th></th>
<th>EQ/CR</th>
<th>2 DOF</th>
<th>CR/PC1</th>
<th>9 DOF</th>
<th>PC1/EQ</th>
<th>6 DOF</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>Test stat</td>
<td>p-value</td>
<td>Test stat</td>
<td>p-value</td>
<td>Test stat</td>
<td>p-value</td>
</tr>
<tr>
<td>KS1</td>
<td>0.073308</td>
<td>0.358</td>
<td>0.029870</td>
<td>0.976</td>
<td>0.050889</td>
<td>0.566</td>
</tr>
<tr>
<td>KS2</td>
<td>0.023478</td>
<td>0.544</td>
<td>0.008464</td>
<td>0.990</td>
<td>0.014332</td>
<td>0.771</td>
</tr>
<tr>
<td>AD1</td>
<td>0.196481</td>
<td>0.338</td>
<td>0.300344</td>
<td>0.021</td>
<td>0.215421</td>
<td>0.088</td>
</tr>
<tr>
<td>AD2</td>
<td>0.054531</td>
<td>0.586</td>
<td>0.025126</td>
<td>0.986</td>
<td>0.038957</td>
<td>0.761</td>
</tr>
<tr>
<td>CvM</td>
<td>0.056979</td>
<td>0.003</td>
<td>0.075237</td>
<td>0.000</td>
<td>0.016837</td>
<td>0.671</td>
</tr>
</tbody>
</table>

**T copula (DOF parameter rounded up from MPL value)**

<table>
<thead>
<tr>
<th></th>
<th>EQ/CR</th>
<th>3 DOF</th>
<th>CR/PC1</th>
<th>10 DOF</th>
<th>PC1/EQ</th>
<th>7 DOF</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>Test stat</td>
<td>p-value</td>
<td>Test stat</td>
<td>p-value</td>
<td>Test stat</td>
<td>p-value</td>
</tr>
<tr>
<td>KS1</td>
<td>0.056672</td>
<td>0.528</td>
<td>0.031377</td>
<td>0.967</td>
<td>0.051822</td>
<td>0.508</td>
</tr>
<tr>
<td>KS2</td>
<td>0.019617</td>
<td>0.568</td>
<td>0.009079</td>
<td>0.980</td>
<td>0.014898</td>
<td>0.726</td>
</tr>
<tr>
<td>AD1</td>
<td>0.211905</td>
<td>0.162</td>
<td>0.301628</td>
<td>0.015</td>
<td>0.209148</td>
<td>0.105</td>
</tr>
<tr>
<td>AD2</td>
<td>0.048571</td>
<td>0.578</td>
<td>0.026579</td>
<td>0.969</td>
<td>0.040153</td>
<td>0.720</td>
</tr>
<tr>
<td>CvM</td>
<td>0.074938</td>
<td>0.000</td>
<td>0.076569</td>
<td>0.000</td>
<td>0.017061</td>
<td>0.660</td>
</tr>
</tbody>
</table>

**Trivariate**

**Gaussian**

<table>
<thead>
<tr>
<th></th>
<th>EQ/CR/PC1</th>
<th>Gaussian</th>
<th>Test stat</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mardia</td>
<td>0.358774</td>
<td>0.228</td>
<td>15.3133</td>
<td>0.337</td>
</tr>
<tr>
<td>KS1</td>
<td>0.076139</td>
<td>0.156</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KS2</td>
<td>0.026872</td>
<td>0.263</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AD1</td>
<td>0.186289</td>
<td>0.218</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>----------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AD2</td>
<td>0.06809</td>
<td>0.248</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CvM</td>
<td>0.085671</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

**T-copula fitted using MPL**

<table>
<thead>
<tr>
<th>EQ/CR/PC1</th>
<th>T DOF = 4.5</th>
<th>Test stat</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>KS1</td>
<td>0.041025</td>
<td>0.811</td>
<td></td>
</tr>
<tr>
<td>KS2</td>
<td>0.016662</td>
<td>0.675</td>
<td></td>
</tr>
<tr>
<td>AD1</td>
<td>0.360425</td>
<td>0.007</td>
<td></td>
</tr>
<tr>
<td>AD2</td>
<td>0.044977</td>
<td>0.664</td>
<td></td>
</tr>
<tr>
<td>CvM</td>
<td>n/a</td>
<td>n/a</td>
<td></td>
</tr>
</tbody>
</table>

**T-copula (DOF parameter rounded down from MPL value)**

<table>
<thead>
<tr>
<th>EQ/CR/PC1</th>
<th>T DOF=4</th>
<th>Test stat</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>KS1</td>
<td>0.04493</td>
<td>0.752</td>
<td></td>
</tr>
<tr>
<td>KS2</td>
<td>0.017448</td>
<td>0.663</td>
<td></td>
</tr>
<tr>
<td>AD1</td>
<td>0.364076</td>
<td>0.016</td>
<td></td>
</tr>
<tr>
<td>AD2</td>
<td>0.046043</td>
<td>0.637</td>
<td></td>
</tr>
<tr>
<td>CvM</td>
<td>0.068843</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

**T copula (DOF parameter rounded up from MPL value)**

<table>
<thead>
<tr>
<th>EQ/CR/PC1</th>
<th>T DOF=5</th>
<th>Test stat</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>KS1</td>
<td>0.043776</td>
<td>0.743</td>
<td></td>
</tr>
<tr>
<td>KS2</td>
<td>0.016226</td>
<td>0.695</td>
<td></td>
</tr>
<tr>
<td>AD1</td>
<td>0.354755</td>
<td>0.008</td>
<td></td>
</tr>
<tr>
<td>AD2</td>
<td>0.04453</td>
<td>0.638</td>
<td></td>
</tr>
<tr>
<td>CvM</td>
<td>0.074332</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>
A.6. Choosing between candidate models

The Gaussian, T and IT copula models can be regarded as a “nested” set of models. Where this is the case, there are two sets of criteria which may be used to form a judgement on which model is preferred. Both tests assess in some sense whether the additional parameters provide a better fit to the data.

A.6.1. Likelihood ratio test

Suppose Model 1 has $p_1$ free parameters and Model 2 has $p_2$ free parameters ($p_2 > p_1$). We wish to assess whether the $(p_2 - p_1)$ additional parameters are statistically significant.

Let $H_0$ be the null hypothesis “$H_0$: data follow Model 1”.

The test is based on the statistic:

$$D = 2 \frac{log L_2}{log L_1} = -2 log L_1 + 2 log L_2$$

Where $L_i$ is the (natural) log likelihood for Model $i$ (=1, 2) estimated at the MLE parameters for that model.

Under the null hypothesis, $D \sim \chi^2_{p_2-p_1}$.

Large values of $D$ indicate that the fitting could be improved by the inclusion of additional parameters.

The null hypothesis is therefore rejected at confidence level $\alpha$ if $D > \chi^2_{p_2-p_1}(\alpha)$, where $\chi^2_{p_1}(\alpha)$ is the $\alpha^{th}$ percentile of a $\chi^2$ distribution.

At the 95% confidence level, the corresponding $\chi^2_{p_2-p_1}$ values are given by:

<table>
<thead>
<tr>
<th>$p_2 - p_1$</th>
<th>$\chi^2_{p_2-p_1}(0.95)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.84</td>
</tr>
<tr>
<td>2</td>
<td>5.99</td>
</tr>
<tr>
<td>3</td>
<td>7.81</td>
</tr>
</tbody>
</table>

A.6.2. Akaike Information Criterion

The Akaike Information Criterion (AIC) is a form of “penalised likelihood”. The maximum likelihood obtained when fitting a model with a greater number of parameters will increase. However, this comes at the risk of over-fitting. The AIC therefore applies a penalty to the additional parameters. The resulting statistic can then be used to form a view on the appropriateness of adding additional parameters.

The AIC is given by the formula

$$AIC = 2p - 2log L$$

Where $L$ is the maximum likelihood and $p$ is the number of additional parameters.

In general, models with lower values of $AIC$ are preferred.

Note that there are other forms of Information Criteria which differ in the penalties they apply to the additional parameters. For example, the Bayesian Information Criterion (BIC) is given by:

$$BIC = p log N - 2log L$$

where $N$ is the number of observations in the sample data. For the purposes of this analysis, we have used the AIC.

Note that the AIC provides a criterion to inform the selection of a model. It does not produce “$p$-values” that would form the basis of a statistical test to decide whether or not to reject a model.
A.6.3. Results

The tables below show the results of applying the Akaike Information Criterion (AIC) and the D statistic from the Likelihood Ratio Test to decide whether the increasing complexity of the model is warranted.

Recall from appendix A.6.2 that we prefer models that result in a lower AIC.

The Likelihood Ratio Test indicates that adding \( p \) additional parameters is significant at the 95% confidence level if \( D \) exceeds \( \chi^2_p(0.95) \).

\[
\begin{array}{c|c}
 p & \chi^2_p(0.95) \\
1 & 3.84 \\
2 & 5.99 \\
3 & 7.81 \\
\end{array}
\]

In each case, the model preferred by the individual criterion is highlighted in blue.

The criteria indicate the following:

Bivariate models

- A copula more complex than Gaussian is warranted only in the case of EQ/CR where a T-copula is preferred by both the AIC and Likelihood Ratio test.

Trivariate models

- The Gaussian copula is rejected in favour of the T-copula (with one extra parameter) under both criteria
- The additional parameters and complexity of the IT copula does not appear to be warranted.

Bivariate

<table>
<thead>
<tr>
<th>EQ/CR</th>
<th>Rho</th>
<th>Nu1</th>
<th>Nu2</th>
<th>logL</th>
<th>AIC</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>-48.8%</td>
<td></td>
<td></td>
<td>27.63</td>
<td>-53.25</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>-46.5%</td>
<td>2.60</td>
<td></td>
<td>35.07</td>
<td>-66.14</td>
<td>14.89</td>
</tr>
<tr>
<td>IT</td>
<td>-46.5%</td>
<td>2.49</td>
<td>2.69</td>
<td>35.08</td>
<td>-64.15</td>
<td>0.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CR/PC1</th>
<th>Rho</th>
<th>Nu1</th>
<th>Nu2</th>
<th>logL</th>
<th>AIC</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>-31.8%</td>
<td></td>
<td></td>
<td>10.67</td>
<td>-19.35</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>-31.2%</td>
<td>9.40</td>
<td></td>
<td>11.33</td>
<td>-18.65</td>
<td>1.30</td>
</tr>
<tr>
<td>IT</td>
<td>-31.2%</td>
<td>5.99</td>
<td>12.82</td>
<td>11.34</td>
<td>-16.68</td>
<td>0.03</td>
</tr>
<tr>
<td>PC1/EQ</td>
<td>Rho</td>
<td>Nu1</td>
<td>Nu2</td>
<td>logL</td>
<td>AIC</td>
<td>D</td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>------</td>
<td>-----</td>
<td>----</td>
</tr>
<tr>
<td>Gaussian</td>
<td>16.8%</td>
<td></td>
<td></td>
<td>2.84</td>
<td>-3.68</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>16.6%</td>
<td>6.08</td>
<td></td>
<td>4.43</td>
<td>-4.86</td>
<td>3.17</td>
</tr>
<tr>
<td>IT</td>
<td>21.0%</td>
<td>41.91</td>
<td>0.59</td>
<td>5.32</td>
<td>-4.64</td>
<td>1.79</td>
</tr>
</tbody>
</table>

### Trivariate

<table>
<thead>
<tr>
<th>Copula</th>
<th>Correlation matrix</th>
<th>Nu1</th>
<th>Nu2</th>
<th>Nu3</th>
<th>logL</th>
<th>AIC</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gauss</td>
<td>100.0%</td>
<td></td>
<td></td>
<td></td>
<td>38.34</td>
<td>-70.68</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-48.8%</td>
<td>100.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>17.1%</td>
<td>-31.8%</td>
<td>100.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>100.0%</td>
<td></td>
<td></td>
<td></td>
<td>4.51</td>
<td>45.53</td>
<td>-83.05</td>
</tr>
<tr>
<td></td>
<td>-48.6%</td>
<td>100.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14.38</td>
</tr>
<tr>
<td></td>
<td>15.0%</td>
<td>-28.7%</td>
<td>100.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>100.0%</td>
<td></td>
<td></td>
<td></td>
<td>2.52</td>
<td>2.91</td>
<td>12.39</td>
</tr>
<tr>
<td></td>
<td>-46.5%</td>
<td>100.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-82.37</td>
</tr>
<tr>
<td></td>
<td>14.1%</td>
<td>-29.3%</td>
<td>100.0%</td>
<td></td>
<td></td>
<td></td>
<td>3.31</td>
</tr>
</tbody>
</table>

### A.7. Conclusions

1. The statistical analysis confirms conclusively that the three risk factors are not independent but does not lead to any firm conclusions on the appropriate form of copula.
2. There is some indication of the presence of tail dependence in the data, particularly between falls in Equity Values and increases in Corporate Bond spreads. However, the majority of our statistical tests do not result in rejection of the Gaussian copula model.
3. Furthermore, the degree of tail dependence appears to vary between pairs of risk factors. If one wished to use a copula which models tail dependence explicitly, this may indicate that a model more complex than a T-copula would be appropriate. Alternatively, adjustments could be made to the parameters of the chosen model.
B. Appendix - Correlations

The following section provides an overview of the various measures of correlation. We give both the population version of the statistic and the final sample version.

B.1. Notation

We use $(X,Y)$ to denote a pair of random variables, $(X_i,Y_i)$ a finite sample from $(X,Y)$ with corresponding rank pairs $(R_i,S_i)$.

B.2. Pearson (linear correlation)

B.2.1. Population Version

$$r(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{E[XY] - E[X]E[Y]}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

B.2.2. Finite Sample Version

$$r_n = \left[ \frac{\sum_{j=1}^{n} (X_j - \bar{X})(Y_j - \bar{Y})}{\sqrt{\sum_{j=1}^{n} (X_j - \bar{X})^2 \sum_{k=1}^{n} (Y_k - \bar{Y})^2}} \right]$$

where $\bar{X} = \frac{1}{n} \sum_{j=1}^{n} X_j$ is the same mean of the $X_j$.

B.3. Spearman rank correlation

B.3.1. Population Version

$$\rho(X,Y) = r(F_X(X), F_Y(Y))$$

i.e. the Spearman rank correlation is the linear correlation applied to the quantiles of the random variables – the linear correlation of the rank statistics.

B.3.2. Finite Sample Version

$$\rho_n = \left[ \frac{\sum_{j=1}^{n} (R_j - \bar{R})(S_j - \bar{S})}{\sqrt{\sum_{j=1}^{n} (R_j - \bar{R})^2 \sum_{k=1}^{n} (S_k - \bar{S})^2}} \right]$$

B.4. Kendall’s tau

B.4.1. Population Version

Let $(\bar{X}, \bar{Y})$ be an independent copy of $(X,Y)$.

The population version of Kendall’s tau is given by:

$$\tau = E[\text{sign}(X - \bar{X})\text{sign}(Y - \bar{Y})]$$

where $\text{sign}(X)$ denotes the sign of $X$. 

110
B.4.2. Finite sample version

The finite sample version is given by:

$$\tau_n = \frac{P_n - Q_n}{\binom{n}{2}} = \frac{4}{n(n-1)} P_n - 1$$

where $P_n$ is the number of concordant pairs and $Q_n$ is the number of discordant pairs. $\binom{n}{2}$ is the number of ways of choosing 2 items from $n$ without replacement = $n(n-1)/2$.

A pair $(X_i, Y_i)$ and $(X_j, Y_j)$ is concordant if $(X_i - X_j)$ and $(Y_i - Y_j)$ have the same signs and discordant if they have different signs. This is illustrated in the diagram below.

Figure 6-3 Concordant versus Discordant Pairs
C. Appendix - Tests for Independence

This section provides an overview of various rank statistic based tests for independence. The tests are all based on making the following null hypothesis:

\( H_0: \text{X and Y are independent} \)

The value of a rank-based statistic is then calculated assuming that the hypothesis is true. The probability of actually observing that value of the statistic is then assessed and the hypothesis rejected if that probability is lower than a defined value or, equivalently, if the value of the statistic exceeds a percentile of its distribution under the hypothesis (that percentile corresponding to some confidence level).

C.1. Spearman test

This test is based on the sample version of Spearman’s rank correlation statistic:

\[
\rho_n = \frac{\sum_{j=1}^{n} (R_j - \bar{R})(S_j - \bar{S})}{\sqrt{\left( \sum_{j=1}^{n} (R_j - \bar{R})^2 \right) \left( \sum_{k=1}^{n} (S_k - \bar{S})^2 \right)}}
\]

Using \( \bar{R} = \frac{1}{n} \sum_{j=1}^{n} R_j \) and the corresponding formula for \( \bar{S} \), the formula above becomes

\[
\rho_n = \frac{12}{n(n+1)(n-1)} \sum_{j=1}^{n} R_j S_j - \frac{3(n+1)}{(n-1)}
\]

Under \( H_0 \), \( \rho_n \) is asymptotically distributed as a Normal random variable \( N \left( 0, \frac{1}{n-1} \right) \)

The test for independence therefore involves calculating the value of \( \rho_n \) and rejecting the null hypothesis at a confidence level of \( \alpha \) if \( \sqrt{\frac{n-1}{(n-1)}} |\rho_n| > z_{\alpha/2} \), where \( z_{\alpha} \) denotes the \( \alpha^{th} \) percentile of the standard Normal distribution.

C.2. Spearman test

This based on the sample version of Kendall’s rank correlation statistic:

\[
\tau_n = \frac{P_n - Q_n}{\binom{n}{2}} = \frac{4}{n(n-1)} P_n - 1
\]

Under \( H_0 \), \( \tau_n \) is asymptotically distributed as a Normal random variable \( N \left( 0, \frac{2(2n+5)}{9n(n-1)} \right) \)

The test for independence therefore involves calculating the value of \( \tau_n \) and rejecting the null hypothesis at a confidence level of \( \alpha \) if \( \sqrt{\frac{9(n-1)}{2(2n+5)}} |\tau_n| > z_{\alpha/2} \).

C.3. Van den Waerden test

This test is based on the Van den Waerden statistic

\[
V_n = \sum_{i=1}^{n} \Phi^{-1} \left( \frac{R_i}{n+1} \right) \Phi^{-1} \left( \frac{S_i}{n+1} \right)
\]

where \( \Phi \) is the cumulative distribution function of the standard Normal distribution.
Under $H_0$, $V_n$ is asymptotically distributed as a Normal random variable $N(0, \sigma^2)$

where $\sigma^2 = \frac{1}{n-1} \left( \sum_{i=1}^{n} \left( \Phi^{-1}\left(\frac{i}{n+1}\right) \right)^2 \right)$

The test for independence therefore involves calculating the value of $V_n$ and rejecting the null hypothesis if $\frac{|V_n|}{\sigma} > z_{\alpha/2}$.

C.4. Cramér-von-Mises test

The bivariate version of the Cramér-von-Mises test is based on a measure of the “distance” between the observed empirical copula and the copula of two independent random variables.

The distance is measured in terms of the usual Euclidean measure of distance:

$$B_n = n \int_0^1 \int_0^1 (C_n(u,v) - uv)^2 du dv$$

where $C_n$ is the empirical copula derived from the sample and the term $uv$ is the copula of 2 independent random variables. $B_n$ is therefore a measure of how far away the empirical copula is from the independence copula over all points.

Under $H_0$, $B_n$ is asymptotically distributed as a sum of independent, identically distributed $\chi_1^2$ random variables.

The test for independence therefore involves calculating the value of $B$ and rejecting the null hypothesis if the probability of observing that value is lower than the required confidence level.

Calculating the probability involves determining the distribution of B. The distribution function of B does not have a simple closed form. In practice, it is more convenient computationally to determine the distribution function using simulation techniques. These involve simulating a large number of samples of n pairs of random variables under the assumption of independence, calculating the statistic B for each simulation and building up the distribution. The observed value of B can then be compared with this empirical distribution and corresponding p-value determined. This test has been implemented by the function “indepTest” in the “copula” package of R.

There is an obvious generalisation of the test to higher dimensions.

Further details may be found in [ref. GR].