Joint defaults in a non-normal world: empirical estimations and suggestions for Basel Accords based on copulas

Fernando F. Moreira

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Management Science and Business Economics

The University of Edinburgh

2010

STATEMENT OF ORIGINAL AUTHORSHIP

This thesis contains no material that has been accepted for the award of any other degree or diploma in any university.

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made.

Fernando F. Moreira

December, 2010

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS, SYMBOLS AND NOTATIONS

BCBS: Basel Committee on Banking Supervision

cdf: cumulative distribution function

GoF: Goodness-of-fit (test)

PD: probability of default

pdf: probability density function

$\forall$: for all

$x \vee y$: max(x, y), that is, the maximum of x and y

$\setminus \{x\}$: except for x (especially when dealing with sets)

ρ : correlation coefficient

ρ_S : Spearman's rho

τ : Kendall's tau

f_X : density function of variable X

F_X : cumulative function of variable X

K_V: Vasicek Formula (gives the probability of default in an extremely adverse scenario)

$\Re$: set of real numbers

ACKNOWLEDGMENTS

I thank the following people and institutions whose contribution helped me throughout my PhD:

My supervisor, Prof. Jonathan Crook, for the fruitful discussions, stimulating questions, incentive and politeness.

The Bank that provided the credit card dataset.

Prof. David Denis and Prof. Mara Faccio (both from Purdue University) for the support.

Secretary of State for Education and Science/United Kingdom (Overseas Research Students Awards Scheme) and Abbey/Santander Bank for financial support for payment of my tuitions.

Credit Research Centre (University of Edinburgh), Midwest Finance Association/US, and the University of Edinburgh Business School for financial support for participation at conferences.

To my father Daltro (*in memoriam*), my mother Marta,

and...

… to those who are brave enough to swim against the stream when nobody else realises that the water is going from the mouth of the river to its source.

"Imagination is far more important than knowledge. (...) The mere formulation of a problem is far more important than its solution which may be merely a matter of mathematical or experimental skills. To raise new questions, new possibilities and to regard old problems from a new angle requires creative imagination and marks real advances in science."

(Albert Einstein)

ABSTRACT

Credit risk models widely used in the financial market nowadays assume that losses are normally distributed and have linear dependence. Nevertheless it is well known that asset returns (loans included) are not normally distributed and present tail dependence. Therefore the traditional approaches are not able to capture possible stronger association among higher losses and tend to underestimate the probability of joint extreme losses.

Copula functions are an alternative to overcome this drawback since they yield accurate dependence measures regardless of the distribution of the variables analysed. This technique was first applied to credit risk in 2000 but the studies in this field have been concentrated on corporate debt and derivatives. We filled this gap in the literature by employing copulas to estimate the dependence among consumer loans. In an empirical study based on a credit card portfolio of a large UK bank, we found evidence that standard models are misspecified as the dependence across default rates in the dataset is seldom expressed by the (Gaussian) copula implicit in those models. The comparison between estimations of joint high default rates from the conventional approach and from the best-fit copulas confirmed the superiority of the latter method.

The initial investigation concerning pairs of credit segments was extended to groups of three segments with the purpose of accounting for potential heterogeneous dependence within the portfolio. To do so, we introduced vine copulas (combinations of bivariate copulas to form high-dimension copulas) to credit risk and the empirical estimations of simultaneous excessive defaults based on this technique were better than both the estimations from the pairwise copulas and from the conventional models. Another contribution of this work concerns the application of copulas to a method derived from the limited credit models: the calculation of the capital required to cover unexpected losses in financial institutions. Two models were proposed and, according to simulations, outperformed the current method (Basel) in most of the scenarios considered.

CHAPTER 1
INTRODUCTION

"Everyone believes in the normal law;
the experimenters because they imagine that it is a mathematical theorem,
and the mathematicians because they think it is an experimental fact."
(Gabriel Lippmann –
quoted by Jules Henri Poincaré in "Calcul des probabilités", 1896)

1.1 PREAMBLE

Credit models widely used in the financial market assume that losses are normally distributed and normally dependent. Due to certain properties of the normal distribution (univariate and multivariate), this presumption makes the calculations easier and accessible to more users (academics and practitioners). Nonetheless the relative simplicity comes at the expense of accuracy and potential extreme losses may be underestimated which can negatively affect research conclusions and the liquidity of financial institutions.

Two aspects are questionable in these models. First, the assumption of normal losses does not seem to be the most adequate. As Bernstein (1996) points out, normally distributed events are typical for natural phenomena but do not represent well facts derived from decisions made by people, such as in the field of economics and finance. Since Mandelbrot (1963), many empirical studies have confirmed this idea and have shown that, in general, financial assets are not normally distributed.

Second, and likely the most important in the context of *portfolio* evaluations, normal dependence (which is implicit in the multivariate normal distribution and can be satisfactorily measured by the linear correlation coefficient) is not able to represent oscillations of returns (or losses) in financial markets where extreme values tend to cluster. Bouyé et al. (2000) emphasise that, even though it is well known that asset returns are fat-tailed, people generally use normal processes to model asset returns because they have more tractable properties for computation. Andrade et al. (2000) show the implications from not considering the dependence of variables that are dependent. Dependence does not affect the mean of the variables but it can very badly affect the variance, which will affect the extreme values of the total distribution. So, in the independence case, it is sufficient to know the two marginal distributions to construct the joint distribution. When the variables are dependent, we also need to know the type and extent of the dependence.

The difficulties mentioned above (namely, assumption of normality and joint distribution modelling) can be solved by using copula functions (explained in Chapter 2, Section 2.3) which can be employed to find the dependence between variables irrespective of their individual distributions. This approach was introduced in finance by Frees and Valdez (1998) and popularised by Embrechts et al. (2002)[1] who explained in detail the benefits of measuring dependence by means of copulas instead of using the linear correlation that is accurate only for some specific distributions (which includes the normal). Chai et al. (2008) present a survey of studies dealing with the application of copulas to finance.

Copulas were first employed in the credit risk field by Li (2000) who corroborated their advantages to model tail relationships. Many posterior investigations have reached the same conclusion (see Chapter 3). Kiesel et al. (2006), for example, state that symmetric return distributions are seldom found empirically and high losses (negative returns) are more likely to occur together than high gains. Asymmetry, which is measured by the distribution's skewness, is especially common for credit risk and the probability of great losses is higher than the Value at Risk (VaR) implied by the normality assumption. The authors show that a positively skewed distribution of losses (negatively skewed returns) results in a skewness-adjusted estimated 99% VaR that is higher than the "expected loss plus 2.33 standard deviations" rule calculated under the assumption of normally-distributed losses.

[1] This study appeared first in 1999 as a working paper.

1.2 MOTIVATION AND AIMS

Like the traditional credit models related to dependence modelling, copulas have not been specifically applied to retail credit. Although consumer loans represent a considerable proportion of banks' assets and have increased significantly in the past decades (Thomas et al., 2005 and Crook et al., 2007), the models typically applied to estimate the dependence in these portfolios were created for corporate debt (Thomas, 2009) and are not compatible with retail credit due to specific characteristics of consumers (Andrade and Thomas, 2007). Furthermore, it is interesting to note that models for *individual* assessments of consumer debtors ("credit scoring") have existed for over 50 years but there is not a specific method established for *portfolios* of consumer debts (Thomas, 2009). Therefore there is a clear necessity of additional research on retail credit at the portfolio level.

One practical consequence of the shortcomings present in conventional credit models is the determination of the capital to be held by financial institutions to face unexpected credit losses (Basel Accords II and III) since the formula used to estimate such capital is derived from methods that presume normally-distributed losses associated through a normal structure (i.e. the dependence is supposed to be adequately measured by the linear correlation coefficient; see Chapter 5 for more details). Given that the formula established in Basel II to estimate the capital to cover extreme credit losses was kept in Basel III, we refer to Basel II as the current Basel Accord in terms of the calculation of the capital needed to cover unexpected credit losses.

Given these gaps in the literature, this thesis aims to: (i) check whether consumer loans also present asymmetric dependence as has been shown for corporate debt and other assets classes, (ii) compare estimations of joint high defaults assuming normality with estimations by means of copulas, (iii) extend the two prior investigations that consider pairs of credit segments to analyses focused on groups of three segments such that we can incorporate possibly different dependence structures (copulas) into one single calculation, and (iv) use Copula Theory to develop more efficient models for Basel Accords with respect to the estimation of tail dependence (which means more precise inferences about coexistent unexpected losses in downturns).

1.3 CONTRIBUTIONS

This thesis makes both empirical and theoretical contributions. The former is associated to the application of Copula Theory to estimate dependence structures across defaults of consumer loans (Chapters 3 and 4) and the latter refers to the suggestion of models to assess unexpected credit losses (Chapters 5 and 6).

The empirical contributions which this thesis makes are as follows. First, we estimate the best-fit copulas for a portfolio of *consumer* loans while the existent studies in the credit area are limited to corporate debt and derivatives. Second, we include five families of copulas that have been rarely or never considered in empirical estimations of dependence in loan portfolios. This is important because the greater number of options increases the possibility of finding copulas that better represent the portfolio studied. Third, we perform goodness-of-fit (GoF) tests limited to the right-tail distribution of the default rates along with the usual tests that consider the whole distributions. This innovation is explained by the fact that the best-fit copulas found by means of the GoF tests will be used to estimate losses in the right tail of the joint distributions (i.e. high default rates). Fourth, we use vine copulas to estimate the dependence across *three* credit segments together (instead of pairs of segments). This technique has not been applied to model dependence in any class of loans or credit derivatives before.

The theoretical contribution of this thesis is to propose the use of copula-based functions in the Basel formulas for capital adequacy. The approaches suggested here are not conditional on the assumption of normality and therefore overcome the limitation of Basel Accords in terms of capturing tail dependence across credit losses. The methods presented in Chapters 5 and 6 can be very attractive for regulators because these formulas can be easily implemented in financial institutions (irrespective of their size) around the world. The most important contribution in this sense is however to *open new avenues* to the development of approaches able to identify potential stronger connection among defaults in downturns. Even if the models are not employed as they are presented here, they can be improved in order to generate more accurate estimations that do not have the same deficiencies present in the methods used nowadays.

Each model has specific strengths. The procedure introduced in Chapter 5 is more flexible and can be adjusted to *any* copula that represents the dependence among credit losses. It can, for example, be differently set for distinct credit classes according to the copula found to be the most representative for the particular type of loan (corporate, mortgage, revolving, etc.).

On the other hand, the principal contributions of the formula derived in Chapter 6 are its compatibility with some negative values of correlation (which is not supported by Basel method) and the relaxation of the assumption of normally-distributed variables.

1.4 IMPORTANCE

The results of this study will benefit many agents involved in the activities of the financial market, such as academics, policy makers (financial regulators, in special), and financial institutions around the world.

The academic community can extend the innovations in the empirical analyses (namely the goodness-of-fit tests based on the upper-tail of default rates' distributions and the inclusion of more families of copulas) to other credit categories or credit card portfolios spanning longer periods. This is a pioneering study in terms of the application of vine copulas to credit risk so there is considerable space for future investigations in this area, especially to cope with different dependence structures in a same portfolio. Researchers can, for instance, empirically estimate best-fit vine copulas for credit classes other than credit cards or look for the vine composition that yields the closest approximation of the likelihood of simultaneous tail losses. Pertaining to the theoretical models suggested as options for Basel II, they can help the improvement of many current credit models that presume multivariate normality. In terms of policy makers, this work can support future versions of Basel Accords able to identify stronger dependence across defaults in downturns so that the regulatory organisations can reduce the risk of underestimation of the capital demanded to cover unexpected credit losses.

The routines for estimating the best-fit copulas and the probability of concurrent extreme defaults can be internally used by financial institutions. The successful performance of the copula-based estimations over the conventional models will assist lenders with effective evaluations of potential losses at the portfolio level. The findings showing that consumer (credit card) loans present asymmetric dependence structures can be used by financial institutions as a guide to allocate capital across debtors that do not tend to cluster in downturns[2]. So, banks will know which combinations of segments are riskier (particularly in severe economic conditions when the financial institutions may suffer higher losses) and will be able to evaluate whether the respective return is consistent with the risk faced. The employment of the vine-copula approach will improve even more the assessment of potential joint losses given that, in practice, portfolios have heterogeneous dependence.

1.5 MAIN FINDINGS

The empirical analysis showed that, like corporate debt and other financial assets, credit card loans present tail dependence. In most of the cases, this relationship is asymmetric and stronger in the right tail which means that the

[2] This allocation could be based, for instance, on segments that express the debtors' characteristics, such as employment status, housing status, income, etc.

Bank studied is exposed to losses higher than those estimated in accordance with traditional credit models that presume normal dependence (equivalent to the Gaussian Copula that denotes symmetric dependence without tail association). The dependence between the distributions of the default rates of two segments was not normal even when the individual distributions were (virtually) normal[3]. Amongst the ten pairs of segments considered, only one had the dependence structure represented by the Gaussian Copula and neither of the segments was normally distributed. This demonstrates that the assumption of multivariate normality (which implies the Gaussian Copula) is not adequate for expressing dependence in credit portfolios and thus can lead to the underestimation of joint extreme losses. To test this hypothesis, we calculated the likelihood of simultaneous high default rates following two approaches: multinormality and the best-fit copulas. The results were then compared to the joint default rates observed in the credit card dataset. The estimations from copulas were more accurate and the better performance was more significant when the copulas were inferred from goodness-of-fit (GoF) tests based on the right tail of the losses' distributions (rather than their complete distributions). Notwithstanding this approach focused on the right tail of the distributions presented higher underestimation indices and are therefore less attractive from a prudential standpoint (compared to calculations that used the entire distributions of the default rates). This finding points to a trade off between better approximations to the default rates and a tendency to underestimate the probability of high losses.

Ten copulas were tested as potential candidates to characterise the link between the pairs of segments in the credit card portfolio studied and seven families were found to be the best representation (three of them have been seldom included in credit risk research). This difference among the dependence within several pairs implies that the definition of an overall dependence structure must combine multiple copulas.

[3] According to the Jarque-Bera test.

We employed vine copulas (cascades of bivariate copulas that forms high-dimension copulas – see Chapter 4, Section 4.2) to express such heterogeneous relationships and illustrate their application by estimating the dependence among *three* credit card segments together (differently from Chapter 3 where the segments were organised in *pairs*). This investigation identified relationships not captured by separate pairwise copulas. Particularly interesting were the cases where vine copulas showed stronger association across high levels of default rates while bivariate copulas were not able to find this type of relationship that denotes the possibility of proportionally higher losses in downturns. Furthermore, the vine-copula approach improved the performance of estimations of concurrent extreme losses based on bivariate copulas (besides outperforming the evaluations that assumed multivariate normality). Another benefit of the vine copulas was the lower level of underestimation of simultaneous high losses when compared with the underestimation from bivariate copulas. This is evidence that the use of vine copulas not only improves estimations of the likelihood of concurrent extreme losses but also reduces the chance of underestimation. The main explanation of this finding is that the credit card portfolio has heterogeneous dependence (different copulas for different pairs) and vine copulas can model such structure more efficiently.

The mathematical expression used to calculate capital needed to avoid failure of financial institutions owing to unexpected credit losses (Basel Accord) comes from traditional credit models and therefore is not able to detect potential stronger dependence among elevated losses (which increases the default rates even more in downturns). As a consequence, the capital required by regulators may be insufficient to compensate the losses in unfavourable periods. The first model suggested (Chapter 5) is suitable for any copula family and thus can be set up according to the relationship empirically found in credit portfolios. The second alternative model (Chapter 6) is grounded in one specific copula and apart from relaxing the assumption of normality it has the advantage of being extendable to some levels of negative correlation (whereas Basel Accord is limited to positive correlation). Simulations revealed that both proposed methods outperformed the Basel approach in most of the scenarios tested.

1.6 THESIS OUTLINE

After this introduction, the thesis is made up of four essays preceded by one review of literature and followed by some conclusions and possible extensions.

Chapter 2 presents a review of literature on topics commonly pertinent to the four essays. It discusses some basic measures of dependence, Copula Theory, dependence modelling via factor credit models, and application of copulas to credit risk studies. Subjects (such as concepts, models and theories) with specific application to one of the essays are in the second section of the respective chapters.

Chapter 3 uses data on a credit card portfolio of a large UK bank to achieve aims (i) and (ii) pointed out in Section 1.2. It first describes some procedures regarding the search for the best-fit copulas. Then the credit card dataset, the criteria for its segmentation and the families of candidate copulas are presented. Next, the best-fit copula for each pair of credit segments is estimated in order to check if the default dependence is asymmetric. Two goodness-of-fit approaches are used: one based on the complete distributions of losses[4] and other based only on their right tail. In the subsequent section, the probability of simultaneous high defaults for each pair of segments is evaluated according to two methods: multivariate normal (equivalent to the traditional models) and the best-fit copulas. The results are compared to the extreme default rates observed in the credit card dataset so that we can decide which approach is more efficient to forecast the likelihood of coincident high losses.

[4] Note that, throughout the thesis, the terms "losses" and "default rates" are used interchangeably. Albeit the data used to estimate the copulas refer to default rates, both expressions are intimately linked since, here, high (low) default rates imply necessarily high (low) losses.

Chapter 4 refers to aim (iii) mentioned in Section 1.2 and deals with the evaluation of dependence among default rates of three credit segments together in the aforementioned credit card portfolio (instead of two segments as done in Chapter 3). The connection across the segments is assessed by means of vine copulas (combinations of bivariate copulas to form higher-dimensional copulas) which are explained in the second section of the chapter. Then this essay basically repeats the steps carried out in the previous chapter with the major objectives of evaluating the dependence structure concerning asymmetry for triplets of segments and verify whether estimations from vine copulas outperform results from bivariate copulas and/or evaluations based on the multivariate normality.

Chapters 5 and 6 focus on aim (iv) cited in Section 1.2 by applying copula functions in two different ways to develop alternative formulas to calculate the capital necessary to cover unexpected credit losses. Chapter 5 introduces Basel Accords and shows how the calculation of the capital required to protect financial institutions against extreme losses in downturns is derived from traditional credit models. Then a model based on copulas is set up with the objective of identifying possible tail dependence that generates losses higher than those inferred from conventional approaches that adopts the multivariate normal distribution as a cornerstone. Simulations are employed to compare the estimations obtained from the copula-based formula with the results of Basel approach.

In Chapter 6, defaults are assumed to be caused by “shocks”. This interpretation leads to the use of copulas related to Poisson processes to derive a model to estimate the probability of conjunct extreme losses. The suggested model is presented and tested with respect to the adequacy of its evaluations of the likelihood of coexistent extreme losses. The model is applied to the Basel context and results in a formula to estimate credit losses in downturns. Simulations are run to check if the alternative formula is able to outperform Basel method.

Chapter 7 summarises the main conclusions of the four essays and proposes further studies that can overcome the limitations of this thesis, fill remaining gaps in this research area or explore new questions raised here.

1.7 NOTATION

Mathematical expressions and variables are presented in italic and in a different type of font compared to the text. As usual, vectors are in boldface. Uppercase and lowercase letters refer to random variables and their realisations, respectively. Only the formulas that are mentioned in posterior parts of the text are numbered (between square brackets). The extensiveness of the software codes (written in Matlab®) made their presentation in the thesis unfeasible but they are available if needed.

1.8 SUMMARY

This chapter introduced the limitations of traditional credit risk models in terms of the calculation of joint losses and gave an overview of this study by presenting its motivation, aims, contributions, importance and main findings. In the next chapter, we will see some concepts related to dependence and credit risk that form the base of the discussions in the subsequent chapters.

CHAPTER 2
DEPENDENCE, COPULAS AND CREDIT RISK

“Things derive their being and nature by mutual dependence and are nothing in themselves” (Nagarjuna)

2.1 INTRODUCTION

The level of losses in a credit portfolio is intimately related to the *joint* behaviour of the loans that form the portfolio. There are many ways of measuring the relationship among variables (credit losses, in this study) and some measures present limitations that may lead to wrong predictions of simultaneous occurrences at specific levels. This chapter presents some measures of dependence and clarifies that the standard measure used in risk management (the linear correlation coefficient) is just one of many concepts of dependence. It also describes how classical credit models deal with dependence to estimate joint losses and how alternative methods have been applied to overcome potential biases in traditional calculations.

In Section 2.2, we present the conventional dependence measure adopted by leading credit risk models along with two other measures (Kendall’s tau and Spearman’s rho) that are relevant in the context of this study. Next, we introduce copula functions and some families that will be used later in the thesis. In Section 2.4, we (i) explain how some popular credit models (factor models) estimate default dependence and (ii) highlight some drawbacks of these methods. Section 2.5 shows how copulas have been applied to credit risk and some existing gaps in the literature that will be addressed throughout the thesis. The last section concludes.

2.2 CORRELATION AND DEPENDENCE

The way assets are associated determines the performance of a pool of investments. Assets with distinct movements reduce potential losses (and profits) in specific periods given that losses from some of them are offset by gains, or smaller losses, from others. There exist many methods to calculate the association between variables (assets' returns, in this case) but, since the publication of Markowitz (1952), most of the financial literature has adopted the concept of (linear) correlation as a proxy for that association[5]. Therefore one specific definition of a relationship between variables, linear correlation, has replaced the broad idea of dependence and the correlation coefficient (Pearson's correlation), typically denoted by ρ, has become the most traditional measure of dependence. For two random variables X_1 and X_2, it is calculated as:

$$\rho(X_1, X_2) = \frac{Cov[X_1, X_2]}{\sqrt{\sigma^2[X_1]\sigma^2[X_2]}}$$

where $Cov[X_1, X_2] = E[X_1 X_2] - E[X_1]E[X_2]$ is the covariance between X_1 and X_2. $\sigma^2[X_1]$ and $\sigma^2[X_2]$ are the variances of X_1 and X_2, respectively.

Despite the diffuse application of ρ, its accuracy as a dependence measure is compromised when the variables are not normally distributed; see Lehmann (1966), Joe (1997), Frees and Valdez (1998), and Li (2000). Moreover Embrechts et al. (2002) point out several drawbacks of this measure. First, it is restricted to linear dependence, so it does not capture other forms of associations. Second, related to the first flaw, a correlation of zero does not imply independence. Third, some correlation levels in the range [-1,1] may not be attainable in the joint distribution. Finally, ρ is not invariant under strictly increasing transformations which means that the correlation between two

[5] Several concepts and properties of dependence measures are detailed in Joe (1997, Chapter 2), Bouyé et al. (2000), McNeil et al. (2005, Chapter 5), Nelsen (2006, Chapter 5), and Genest and Favre (2007).

variables (X_1 and X_2, for example) is different from the correlation of transformations of those same variables, e.g. the logarithmic expressions $\ln(X_1)$ and $\ln(X_2)$.

On the other hand, rank association measures do not suffer from these problems and, therefore, are more useful in the copula approach (to be explained in the next section) that demands association measures independent of marginal (individual) distributions. Two of those measures are presented below: Kendall's tau (τ) and Spearman's rho (ρ_S).

Kendall's tau is based on the number of concordant and discordant pairs of variables. Assuming (X_1, Y_1) and (X_2, Y_2) are two independent pairs from a joint distribution, they will be *concordant* if $(X_2 - X_1)(Y_2 - Y_1) > 0$, i.e., if the two variables move in the same direction. They will be *discordant* when $(X_2 - X_1)(Y_2 - Y_1) < 0$.

Defining c as the number of concordant pairs and d as the number of discordant ones, Kendall's tau is given by:

$$\tau = \frac{c-d}{c+d}$$

or, equivalently, τ = *Pr[concordance] – Pr[discordance]*, which is

$$\tau = \Pr[(X_2 - X_1)(Y_2 - Y_1) > 0] - \Pr[(X_2 - X_1)(Y_2 - Y_1) < 0]$$

As for the Spearman's rho, if X_1 and X_2 are two random variables with marginal cumulative distribution functions F_1 and F_2, respectively, ρ_S will be the correlation between $F_1(X_1)$ and $F_2(X_2)$. In the multivariate case:

$$\rho_S(\mathbf{X}) = Corr(F_1(X_1),...,F_d(X_d))$$

where $Corr$ refers to the correlation matrix. That is, this measure is a linear correlation among the distribution functions of the variables studied.

2.3 COPULAS

2.3.1 Basic concepts and properties of copulas

Copulas are multivariate distribution functions with uniformly distributed margins in (0,1) that link marginal (individual) distributions of variables to their joint distributions:

$$F_{1...d}(x_1,..., x_d) = C(F_1(x_1),..., F_d(x_d))$$

where $F(.)$ denotes a cumulative distribution function and C stands for a copula. Thus, C is an expression (function) with d inputs and, when evaluated at $F_1(x_1),..., F_d(x_d)$, returns the joint cumulative distribution of the d variables evaluated at $x_1,...,x_d$, i.e., the probability that all variables $X_1,...,X_d$ are concurrently below the respective values $x_1,...,x_d$.

The name "copula" comes for the Latin word "copulare" which means "to couple", "to join" or "to connect" and was chosen to emphasise the manner a copula "couples" a joint distribution function to the variables' univariate margins. A detailed view on copulas is given in Joe (1997) and Nelsen (2006)[6].

Fermanian and Scaillet (2005) and Schmidt (2007) explain that all marginal distributions are transformed into uniform ones so that all variables get the same type of distribution. Hence the intuition behind the copula idea is that the equally-distributed marginals (after the transformation) are used as the

[6] Apart from these classical textbooks, some papers also provide an introduction to copulas, e.g. Frees and Valdez (1998), Quesada-Molina (2003), Trivedi and Zimmer (2005), and Genest and Favre (2007).

reference case and the copulas express the dependence structure according to this reference.

For $1 \le i \le d$, $F_i(x_i)$, the univariate function of a variable X_i, transforms the value x_i into its correspondent percentile (rank, commonly represented by "u" or "v" in the literature), i.e. X_i becomes uniformly distributed in the interval (0,1). Such transformation is explained by the "Probability Integral Transformation" (PIT). Consider a random variable X_i, with continuous cumulative distribution function F_i. The application of F_i to a specific value of X_i, x_i, generates a uniform variable between 0 and 1. That is[7], $F_i(x_i) \sim U(0,1)$.

However such transformations are done by the sake of simplicity. Frees and Valdez (1998) state that the marginal distributions can be of any type. As a matter of fact, although the concept of a copula was formally published in 1959, some earlier studies had already divulged very similar ideas with distributions standardised in different intervals (see Nelsen, 2006, Chapter 1).

Rosenberg and Schuermann (2006) clarify that the shape of the dependence between the variables (e.g. lower/upper tail association, symmetry/asymmetry) is determined by the copula, while the scale and the shape of each variable's distribution (i.e. parameters such as mean and standard deviation in the case of the normal distribution) are completely determined by the marginals.

For the case of two variables, Rockinger and Joundeau (2001) present an intuitive view of a bivariate copula as a function $C:[0,1]^2 \to [0,1]$ with three properties[8]:

1. $C(u,v)$ is increasing in u and v;
2. $C(0,v) = C(u,0) = 0$, $C(1,v) = v$ and $C(u,1) = u$;

[7] The proof of PIT is given, for example, in Casella and Berger (2002, pp. 54-55).

[8] These properties are also valid for *n*-dimensional copulas $C:[0,1]^n \to [0,1]$; see, for instance, Nelsen (2006).

3. For all u_1, u_2, v_1, v_2 in [0,1], such that $u_1 < u_2$ and $v_1 < v_2$ we have $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$.

where u and v represent the percentiles of the variables.

Property 1 means that when one marginal distribution is constant, the joint probability will increase if the other marginal distribution increases.

Property 2 reveals expected conditions for joint distributions: if one margin has zero probability the joint occurrence also has zero probability to occur. Consequently, if on the contrary one margin is certain to occur, then the probability of a joint distribution is determined by the remaining margin probability.

Property 3 states that if the percentiles u and v increase, then their joint probability function also increases. This property is therefore a multivariate extension of the condition that a cumulative distribution function must be increasing.

An informal definition of copula is given in Kolev et al. (2006): let $X_1,...,X_d$ be continuous random variables with distribution function $F_{1...d}(x_1,..., x_d) \in [0,1]$ and marginal distributions $F_1(x_1),..., F_d(x_d)$, each $F(.) \in [0,1]$. The associated *d-dimensional copula* that links all margins $F_1(x_1),..., F_d(x_d)$ to the joint distribution $F_{1...d}(x_1,..., x_d)$ is the mapping from $[0,1]^d$ to $[0,1]$.

A central idea in the copula approach is Sklar's Theorem (due to Sklar, 1959) which provides the foundation of many applications of Copula Theory to practical problems.

Sklar's Theorem: *Let F_{12} be a joint distribution function with margins F_1 and F_2. Then there exists a copula C such that for all x_1 and x_2 in $\Re$,*

$$F_{12}(x_1, x_2) = C(F_1(x_1), F_2(x_2))$$

If F_1 and F_2 are continuous, C is unique. Conversely, if C is a copula and F_1 and F_2 are distribution functions, then the function F_{12} is a joint distribution function with margins F_1 and F_2. This is extendable to higher dimensions.

This finding is essential for the empirical analyses in Chapters 3 and 4 where we will estimate the best-fit copulas concerning a dataset of default rates. Sklar's Theorem implies that, in a portfolio of loans, each group (pair or triplet) of credit segments[9] has only one underlying ("true") copula and, consequently, we know that we are searching for a unique solution.

Whilst the linear correlation coefficient ρ is accurate only for spherically or elliptically distributed data[10] (see Embrechts *et al.*, 2002), copulas are suitable for any type of distribution since they are based on ranks (percentiles).

Nelsen (2006) points out other advantages of using copulas. They are a flexible tool to model dependence (various copulas represent different dependence structures between variables and they allow us to separately model the marginal behaviour and the dependence structure), they indicate not only the degree of the dependence but also the structure of the dependence (they can, for instance, directly model the tail dependence), and copula functions are invariant to transformations of the underlying variables while the correlation is not (i.e. the same copula function can be used, e.g., for both the returns of assets and their logarithm). Moreover copulas do not require normality of the variables studied, which is useful when dealing with dependence between asset returns (especially with high frequency data).

Being multivariate cumulative distribution functions, copulas give the probability of variables being simultaneously below particular values. $C(F_1(x_1), F_2(x_2))$, for

[9] The credit card portfolio was segmented according to the loans' credit quality. The criteria for the segmentation will be presented in Chapters 3 and 4.

[10] For a technical concept of spherical and elliptical distributions, see item 3.3 of Embrechts et al. (2002). Intuitively, in the trivariate case, we can identify such distributions through their contour diagrams (graphs of level curves) which have spherical and elliptical shapes respectively. The normal distribution is an example of this class.

example, is equivalent to the probability that X_1 is smaller than x_1 at the same time that X_2 is smaller than x_2.

It is also possible to use copulas in order to calculate the probability that variables will be jointly *above* specific points. These are the so-called Survival Copulas and have the form (see Nelsen, 2006):

$$\bar{F}_{1...d}(x_1,..., x_d) = \hat{C}(\bar{F}_1(x_1),..., \bar{F}_d(x_d))$$

where $\bar{F}_{1...d}(x_1,..., x_d)$ is the joint probability $\Pr(X_1 > x_1,..., X_d > x_d)$ and, for $1 \leq i \leq d$, each $\bar{F}_i(x_i)$ is a survival (or reliability) function $\Pr(X_i > x_i) = 1 - F(x_i)$. Survival Copulas turn out to be very useful in the context of this study given that the estimation of joint *high* default rates (credit losses) is a common objective in the following four chapters.

The strength of the dependence (copula) is expressed by a parameter θ which is closely related to the rank correlations Kendall's tau (τ) and Spearman's rho (ρ_S) defined in Section 2.2[11]. Let u_1 and u_2 represent $F_1(x_1)$ and $F_2(x_2)$ respectively, the univariate cumulative distribution functions (*cdfs*) of the random variables X_1 and X_2 evaluated at x_1 and x_2. The intensity θ of their representative copula can then be inferred from[12]:

$$\tau = 4\iint_{[0,1]^2} C(u_1,u_2)dC(u_1,u_2) - 1 \qquad [\ 2.1\ ]$$

and

$$\rho_S = 12\iint_{[0,1]^2} C(u_1,u_2)du_1du_2 - 3 \qquad [\ 2.2\ ]$$

[11] This refers to one-parameter copulas. There are also copulas with more parameters but, for the sake of simplicity, these families are not considered in this thesis.
[12] The proofs are given in Nelsen (2006, Chapter 5).

Some relationships between rank correlations and the parameters of specific copulas are given, for example, in Frees and Valdez (1998) and Nelsen (2006). When the closed form exists, this is a faster and relatively easy way (when compared to the maximum likelihood methods presented in Chapter 3) of calculating the parameter of a copula based on the rank correlation between variables.

2.3.2 Families of copulas

There are “families” of copulas in the same way that there are types of distributions. Many families of copulas are described in Joe (1997, Chapter 5) and Nelsen (2006, Chapter 4). We will focus here on the copulas that will be used in the empirical analyses (Chapters 3 and 4). They will be split into three groups according to the classes which they belong to: Elliptical Copulas, Archimedean Copulas and other classes.

These copulas were selected among the bivariate one-parameter families presented in Joe (1997, Chapter 5). The Student t Copula was added to the list since this family has been applied in many financial studies and indicates stronger dependence in the tails of the distributions than the dependence represented by the Gaussian Copula (which is interesting in this context). Two other families (Marshall-Olkin and Cuadras-Augé) have a specific application in Chapter 6 and will be introduced there.

For convenience, in the following formulas, exponential terms with base $e = 2.71828$ (Euler’s number) and exponent a are written as e^a if a has no more than two elements and $\exp\{a\}$ otherwise.

2.3.2.1 – Elliptical copulas

These copulas express symmetric dependence and the contour plots have elliptical shape. Amongst the copulas considered in this study, two belong to this class: the Gaussian (Normal) Copula and the Student t Copula (see Nelsen, 2006, for instance).

- *Gaussian Copula* (symmetric without tail dependence)

This family has been of great interest for credit risk analyses, particularly when the margins are normally distributed (see Section 2.5). In these conditions, it has the expression:

$$C_{Ga}(u_1,..., u_d) = \Phi_{\Sigma}(\Phi^{-1}(u_1),..., \Phi^{-1}(u_d)) \quad [2.3]$$

where Φ_{Σ} is the standard multivariate cumulative normal distribution with linear correlation matrix Σ and Φ^{-1} is the inverse of the standard univariate normal distribution. In the bivariate case, the copula has the following form:

$$\Pr(X_1 < x_1, X_2 < x_2) = C_{Ga}(u_1, u_2; \theta) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{\sqrt{2\pi(1-\theta^2)}} \exp\left\{ \frac{-(r^2 - 2\theta rs + s^2)}{2(1-\theta^2)} \right\} drds$$

where $\theta \in [-1,1]$ and in the case of the Gaussian Copula with normal margins is the linear correlation coefficient between the two random variables X_1 and X_2; $r = \Phi^{-1}(u_1), s = \Phi^{-1}(u_2)$.

- *Student t Copula* (symmetric with tail dependence)

In the usual case of t-distributed marginals, this family is represented by:

$$C_t(u_1,..., u_d) = t_{\nu}^{d}(t_{\nu}^{-1}(u_1),..., t_{\nu}^{-1}(u_d))$$

where ν represents the degree of freedom of the distribution, t_ν^d denotes the standardised multivariate Student t distribution function with d variables and t_ν^{-1} is the inverse of the marginal Student t distribution. For $d = 2$, this copula becomes:

$$\Pr(X_1 < x_1, X_2 < x_2) = C_t(u_1, u_2; \theta, v) =$$
$$\int_{-\infty}^{t_\nu^{-1}(u_1)} \int_{-\infty}^{t_\nu^{-1}(u_2)} \frac{\Gamma\left(\frac{v+2}{2}\right)}{\pi v \sqrt{(1-\theta^2)}\Gamma(v/2)} \left(1 + \frac{1}{v}\mathbf{x}^T (1-\theta^2)^{-1}\mathbf{x}\right) dx_1 dx_2$$

where $\mathbf{x}$ is a vector composed by the cut-off values (x_1, x_2) and $\mathbf{x}^{\boldsymbol{T}}$ is its transpose. Γ denotes the Gamma distribution and $\theta \in [-1,1]$ is the copula parameter, in this case, equivalent to the linear correlation between the variables (see Cherubini et al., 2004).

2.3.2.2 Archimedean Copulas

One class of copulas with great application is the Archimedean class. These copulas have the general form[13]:

$$C(u_1, \ldots, u_d) = \Psi^{-1}[\Psi(u_1) + \ldots + \Psi(u_d)]$$

where $0 \le u_1, \ldots, u_d \le 1$ and Ψ is a function called generator that satisfies the properties:

(i) $\Psi(1) = 0$;

(ii) for all $g \in (0,1)$, $\Psi'(g) < 0$, i.e. Ψ is decreasing; and

(iii) for all $g \in (0,1)$, $\Psi''(g) \ge 0$, i.e. Ψ is convex.

[13] Archimedean copulas are detailed in Nelsen (2006, Chapter 4).

Four bivariate Archimedean copulas will be considered in this study:

- *Clayton Copula* (lower-tail dependence)

Generator $\Psi(g) = g^{-\theta} - 1$ and

$$\Pr(X_1 < x_1, X_2 < x_2) = C_C(u_1, u_2; \theta) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}$$

where $\theta \in [-1, \infty) \setminus \{0\}$.

- *Frank Copula* (symmetric without tail dependence)

Generator $\Psi(g) = -\ln \frac{e^{-\theta g} - 1}{e^{-\theta} - 1}$ and

$$\Pr(X_1 < x_1, X_2 < x_2) = C_F(u_1, u_2; \theta) = -\theta^{-1} \ln\left(\frac{(1 - e^{-\theta}) - (1 - e^{-\theta u_1})(1 - e^{-\theta u_2})}{1 - e^{-\theta}} \right)$$

where $\theta \in (-\infty, \infty) \setminus \{0\}$.

- *Gumbel Copula* (upper-tail dependence)

Generator $\Psi(g) = (-\ln g)^{\theta}$ and

$$\Pr(X_1 < x_1, X_2 < x_2) = C_{Gu}(u_1, u_2; \theta) = \exp\{-[(-\ln(u_1)^{\theta} - \ln(u_2)^{\theta}])^{1/\theta}\}$$

where $\theta \in [1, \infty)$.

- *Joe Copula* (upper-tail dependence)

Generator $\Psi(g) = -\ln[1 - (1-g)^{\theta}]$ and

$$\Pr(X_1 < x_1, X_2 < x_2) = C_J(u, v; \theta) = 1 - [(1-u)^{\theta} + (1-v)^{\theta} - (1-u)^{\theta}(1-v)^{\theta}]^{1/\theta}$$

where $\theta \in [1, \infty)$.

2.3.2.3 Copulas of other classes

The other copulas employed in the empirical investigation are neither Elliptical nor Archimedean (see Joe, 1996, for more details).

- *Farlie-Gumbel-Morgenstern Copula* (symmetric without tail dependence)

$$\Pr(X_1 < x_1, X_2 < x_2) = C_{FGM}(u_1, u_2; \theta) = u_1 u_2 (1 + \theta(1 - u_1)(1 - u_2))$$

where $\theta \in [-1,1]$.

- *Galambos Copula* (upper-tail dependence)

$$\Pr(X_1 < x_1, X_2 < x_2) = C_{Gb}(u_1, u_2; \theta) = u_1 u_2 \exp\{(-\ln(u_1)^{-\theta} - \ln(u_2)^{-\theta})^{-1/\theta}\}$$

where $\theta \in (0, \infty)$.

- *Hüsler-Reiss Copula* (upper-tail dependence)

$$\Pr(X_1 < x_1, X_2 < x_2) = C_{HR}(u_1, u_2; \theta) =$$

$$\exp\left\{\ln(u_1)\Phi\left(\theta^{-1} + \frac{1}{2}\theta\ln\left[\frac{\ln(u_1)}{\ln(u_2)}\right]\right) + \ln(u_2)\Phi\left(\theta^{-1} + \frac{1}{2}\theta\ln\left[\frac{\ln(u_1)}{\ln(u_2)}\right]\right)\right\}$$

where $\theta \in [0, \infty)$.

- *Plackett Copula* (symmetric without tail dependence)

$$\Pr(X_1 < x_1, X_2 < x_2) = C_P(u_1, u_2; \theta) =$$

$$\frac{1 + (\theta - 1)(u_1 + u_2)}{2(\theta - 1)} - \frac{\sqrt{[1 + (\theta - 1)(u_1 + u_2)]^2 - 4u_1 u_2 \theta(1 - \theta)}}{2(\theta - 1)}$$

where $\theta \in [0, \infty)$.

2.4 MODELLING DEFAULT DEPENDENCE (CORRELATION) VIA FACTOR MODELS

Some leading industry credit risk models, such as CreditMetrics® and KMV®, rely on the presumptions of structural models (initially proposed by Merton, 1974) according to which an obligor defaults when a latent variable associated to it (typically interpreted as the log-returns of its assets) falls below a threshold (the amount needed to pay the outstanding debt).

The dependence across defaults of different obligors is estimated in line with factor models which assume that the correlation among defaults is driven by the debtors' latent variables (see, for instance, Crouhy et al., 2000 and Bluhm et al., 2002). Such underlying variables are impacted by common (systematic) factors that affect all obligors and specific (idiosyncratic) factors that have effect only on the respective borrowers.

The idiosyncratic factors are assumed to be independent from one another and therefore do not contribute to asset return correlations which are exclusively determined by the systematic factors.

To illustrate this idea, consider a case based on an example given by Bluhm et al. (2002). If two automotive companies A and B operating in country C are debtors, the ability of those firms to pay their obligations is likely affected *in the same direction* by the underlying factor *automotive industry*. That is, if the activity in that sector falls, the default probability of A and B increases simultaneously. Another aspect that influences the performance of those companies is the country C's economic level. So this is another systematic factor that may change the default probability of A and B in the same way. In contrast, if the firm A's CEO steps down or one of its factories is flooded, this event will, in principle, impact only the default likelihood of A (not B's). Hence, this would be an idiosyncratic risk of A.

Naturally, there are many common factors that act together and influence debtors' situations. However this model may be simplified if we consider that the asset returns of all borrowers are driven by only one common factor (the “economic status”). The latent variable (Y), the single systematic factor (E), and the specific factor (ε) are assumed to be standardized normally distributed. Also, each idiosyncratic risk is uncorrelated with the systematic risk and the specific risks of all other obligors. For simplicity, all pairs of asset returns are considered to present the same correlation (ρ).

Owen and Steck (1962) show that equally correlated and jointly standard normal variables may be expressed as a function of their correlation coefficient and two other standard normal variables. Thus, considering all assumptions of factor models, in the case of a single common risk, the latent variable Y for a debtor i may be expressed as a function of E, ε, and ρ, namely:

$$y_i = E\sqrt{\rho} + \varepsilon_i\sqrt{1-\rho} \qquad [\ 2.4\ ]$$

where $\sqrt{\rho}$ and $\sqrt{1-\rho}$ indicate how much of the variability of y_i is explained by E and ε_i, respectively.

Apart from the doubtful presumption of normal behaviour for some of these variables, the use of the linear correlation coefficient is a limitation given that it does not capture asymmetric dependence which could indicate more or less intense association across some variables in certain scenarios (see Embrechts et al., 2002). Furthermore, note that the correlation between the latent variables, ρ, is not permitted to take negative values and thus we cannot express such variables when they are negatively related (which implies negative correlation between defaults).

As we will show in Chapter 5, [2.4] is used to derive the formula applied to estimate the capital to be required from financial institutions to cover unexpected credit losses. Thus all limitations cited above are reflected in the calculation of capital banks should set aside to guarantee their operations over downturns. Therefore, since the shortcomings of [2.4] may result in an underestimation of the likelihood of joint high defaults, the capital determination is subject to the same drawback that may cause insufficiency of reserves.

2.5 APPLICATION OF COPULAS TO CREDIT RISK

To the best of our knowledge, the first application of copulas to credit risk analyses is due to Li (2000) who showed that although factor credit models[14] do not employ the concept of copulas explicitly, the estimation of joint default probability following those methods corresponds to the use of a bivariate Gaussian Copula whose parameter is expressed by the linear correlation between default probabilities within pairs of loans.

Let Y, a standard normal random variable, be a latent variable that guides loans' default. Considering two debts A and B, the default of A and B will happen in a particular period (one year, for example) if the latent variable falls below the cutoffs y_A and y_B respectively in that time horizon. Thus, the probability of default for A, pd_A, and B, pd_B, in the analysed period is:

$$pd_A = \Pr[Y < y_A] \text{ and } pd_B = \Pr[Y < y_B] \qquad [2.5]$$

Denote the asset returns' correlation that is used as proxy for the correlation between the latent variables as ρ. In factor models, the conjunct default probability of A and B is obtained from:

$$\Pr[Y < y_A, Y < y_B] = \int_{-\infty}^{y_A}\int_{-\infty}^{y_B} \phi(x_A, x_B; \rho)dx_A dx_B = \Phi(y_A, y_B; \rho) \qquad [2.6]$$

[14] The author mentions CreditMetrics® but his reasoning may be extended to other factor credit models such as KMV®.

where $\phi(x_A, x_B; \rho)$ is the standard bivariate normal density function of asset returns x_A and x_B with correlation ρ, and Φ stands for the bivariate normal cumulative distribution function.

Assume now that we want to estimate the probability of default within one year. Let T_A and T_B represent the survival times of debts A and B. Hence, the joint probability of default in that period is given by $\Pr[T_A < 1, T_B < 1]$ and the individual probability of default over the same period is:

$$pd_i = \Pr[T_i < 1] = F_i(1) \qquad [2.7]$$

for any debt i and distribution function $F_i(.)$ of the survival times evaluated at one year.

From [2.5] and recalling that the latent variable Y follows a standard normal distribution, we know that, for any debt i:

$$Y_i = \Phi^{-1}(pd_i) \qquad [2.8]$$

Combining [2.7] and [2.8], we get:

$$Y_i = \Phi^{-1}(F_i(1)) \qquad [2.9]$$

Assuming that T_A and T_B have correlation parameter γ and using [2.9], another way of calculating [2.6] is:

$$\begin{gathered}\Pr[Y < y_A, Y < y_B] = \Phi(y_A, y_B; \rho) = \\ \Pr[T_A < 1, T_B < 1] = \Phi(\Phi^{-1}(F_A(1)), \Phi^{-1}(F_B(1)); \gamma)\end{gathered} \qquad [2.10]$$

If we set $\rho = \gamma$ (i.e. the correlation between the latent variables is equal to the correlation between the survival times), we can see that the last term in [2.10], $\Phi(\Phi^{-1}(F_A(1)), \Phi^{-1}(F_B(1)); \gamma = \rho)$, corresponds to a Gaussian Copula (with parameter ρ) mentioned in Section 2.3.2.1 (see expression [2.3] where each u represents a variable uniformly distributed in [0,1], the same feature of $F(.)$ in [2.10]). The distribution functions of the survival times are evaluated at one year in this example but, obviously, any time horizon could be inserted in the expression and $F_i(.)$ would be bounded at [0,1].

After Li (2000), a number of papers have shown the benefits of using joint probability of default without the assumptions of factor credit models concerning the normality of variables and the Gaussian dependence.

Frey et al. (2001), Frey and McNeil (2001, 2003), Bluhm et al. (2002, Chapter 2), and Kostadinov (2005) conducted simulation studies and found that the assumption of normality of the latent variables can have an impact on the distribution of credit losses such that joint defaults may be incorrectly estimated. In particular, they found that the use of the Gaussian Copula (as implicit in well known credit models) resulted in the underestimation of credit exposure while the Student t Copula yielded better evaluations (which is consistent with the evidence from the financial literature where asset portfolios have been found to have more mass in the tails than presumed by the normal distribution). Bo-Chih (2004) used data on corporate obligors of Taiwanese banks to compare those two copulas with respect to the identification of potential extreme losses and confirmed the superiority of the Student t Copula. Schmidt (2003) showed that elliptical copulas (the Gaussian excluded) have properties related to tail association and therefore can incorporate the dependence structure of extreme default occurrences. The outperformance of these copulas was limited to predictions concerning defaults driven by a unique risk factor (systematic risk). Kang and Shahabuddin (2005) suggested some procedures to incorporate multiple t-distributed factors in estimations of conjunct extreme defaults. A more flexible way of using the multivariate Student t Copula was proposed by Daul et al. (2003) in order to identify more precisely the association among different types of risk factors. The resultant structure, called Grouped t Copula, has margins with different shapes and is useful for capturing the dependence among variables with unlike tail dependence. This copula was found by Di Clemente and Romano (2004) to be the best option (compared to the Gaussian, the Clayton, and the Student t) to model the tail dependence in a loan portfolio composed of ten Italian (corporate) obligors.

The only exception for these conclusions against the suitability of the Gaussian Copula is Hamerle and Rösch (2005) who analysed the performance of credit loss forecasts by using the Gaussian Copula when the (true) underlying dependence comes from the Student t Copula. The authors used corporate debt data from Standard & Poors and found that, although the asset correlations are biased, the losses (Value-at-Risk) forecasted via the Gaussian Copula are alike to the “true” losses.

Apart from the elliptical copulas, other families have been tested in credit studies. Hamilton et al. (2001), for instance, adopted the Ait-Mikhail-Haq Copula to express dependence between default rates of different classes of bonds. The authors did not test other families but showed that their choice was better than the multivariate normal distribution to represent that dependence. Schönbucher and Schubert (2001) suggested a model based on copulas to associate continuously updated probabilities of default. The approach was illustrated with three families (Gaussian, Clayton and Gumbel) and was found to yield realistic distributions of default times. Melchiori (2003) used Archimedean copulas (Clayton, Frank and Gumbel) to price first-to-default contracts. He concluded that Gumbel was the best choice to deal with the dependence in his example. Das and Geng (2006) simulated the dependence of US corporate credit by first estimating the margins and then estimating the copula. Their best results were obtained from the combination of skewed double exponential marginals with the Clayton Copula. Cherubini et al. (2004) and Hull and White (2004, 2006) presented examples of the application of copulas (Student t, Clayton and Frank) to model default correlation in credit derivatives such as CDOs (Collateralized Debt Obligations) and swaps. All three studies confirmed that the Gaussian Copula was not the best representation for most of the empirical data analysed.

2.6 CONCLUSIONS

The linear correlation coefficient does not yield good representation of the relationship between non-normal variables. In contrast, some dependence measures, such as Kendall's tau and Spearman's rho, do not have this limitation and are compatible with the use of copulas to estimate joint occurrences of variables regardless of their individual distributions.

Different families of copulas capture distinct dependence structures and traditional credit models (that use the linear correlation) implicitly adopt a dependence shape given by the Gaussian (Normal) Copula which, in turn, means that these models are subject to the misspecification of joint occurrences (especially the extreme ones) when losses are not normally distributed. As we mention in the next chapter, there is evidence that credit losses do not follow the normal distribution. Thus we can infer that estimates of joint losses based on copulas must be more precise.

The use of copulas in credit risk studies has been concentrated on corporate debt and credit derivatives and the most frequently considered classes have been the elliptical (mainly Gaussian and Student t) and the Archimedean (specially Clayton, Frank, and Gumbel). This thesis will extend the use of copulas in credit risk management by applying them to consumer loans and by testing some families seldom considered in this research field, namely: Farlie-Gumbel-Morgenstern (FGM), Galambos, Hüsler-Reiss, Joe, and Plackett.

CHAPTER 3
CHECKING FOR ASYMMETRIC DEFAULT DEPENDENCE IN A CREDIT CARD PORTFOLIO: A COPULA APPROACH

"I can't understand why people are frightened of new ideas.
I'm frightened of the old ones." (John Cage)

3.1 INTRODUCTION

Since the 1960's there is abundant evidence in the literature showing that asset returns in general are not normally distributed (see Mandelbrot,1963 and Fama,1965) and many empirical studies have confirmed this behaviour for several classes of investments, including loan portfolios (Rosenberg and Schuermann, 2006).

Moreover, it has also been found that returns are more correlated in the left tail (i.e. when investments result in losses or lower returns) than in the right tail. See, for instance, Ning (2010), who cites many other studies that reach the same conclusion, Ang and Bekaert (2002), and Patton (2006). According to Di Clemente and Romano (2004) and Das and Geng (2006), returns of credit assets also present asymmetric (tail) dependence.

Factor credit risk models assume that returns of obligors' assets are normally distributed, not only individually (univariate normal distribution for each debtor's asset returns) but also at the portfolio level (joint distribution of asset returns represented by the multivariate normal). This implies relatively fewer occurrences of simultaneous extreme values than if appropriate distributions were used and therefore may lead to biased estimations if asset returns do not follow that particular distribution.

The first contribution of this chapter is the empirical estimation of best-fit copulas for consumer loans by using a credit card dataset provided by a large UK bank. We estimate *joint* extreme default rates based on copulas and compare them to estimations conditional on the assumption of normality. No previous studies have performed such estimates for a credit card portfolio.

The second contribution is to test five copulas that are not usually included in research pertaining to credit risk. As a consequence, we will be able to infer the applicability to credit card portfolios of different copulas. For practitioners, this will also help to improve the estimation of joint credit losses inasmuch as the choice of copulas in financial institutions is usually arbitrary or justified by convenience and tractability (Student t Copula, for instance, is often easy to simulate by using Monte Carlo method); see Jouanin et al. (2004).

A third innovation is the use of goodness-of-fit tests (GoF) based only on the *right tail* of the variables' distribution (as opposed to usual procedures that consider whole distributions). This strategy was implemented because the principal objective of finding the best-fit copulas here is to employ them to estimate the probability of joint *high* defaults.

A sample of credit card loans was split into five segments according to a score provided by the Bank. Then the association between each of the ten pairs of segments was modelled by the best-fit copula. Most of the pairs of segments present right-tail dependence which suggests the existence of flaws in estimations of joint high defaults derived from traditional (factor) models (which do not detect stronger connections across extreme values of the variables). In other words, such structure means that higher default rates are more associated and the Bank is subject to larger losses in downturns than would be calculated with traditional techniques. We also find that some of the pairs have dependence represented by three of the five *less popular* copulas inserted in this study.

After finding the best representation for the dependence across the credit card loans, we compare estimations of conjunct "high" default rates following conventional assumptions of normality and Copula Theory. In most cases, the latter method generates values closer to the observed default rates in the dataset. Considering each pair of segments separately and six risk levels (loss percentiles), the copula approach gave overall better results for all pairs.

The remainder of this chapter is organised as follows. Section 3.2 contains a brief review of techniques to estimate copula parameters and of goodness-of-fit tests (GoF) to decide which copula is the best one among many candidates. Next, we describe the data used in the empirical analysis. The ten copula families taken as candidates to represent the dependence structure among credit card loans are introduced in Section 3.4. Then, we estimate the dependence structure (copulas) between pairs of segments in a credit card portfolio of a large UK bank. Section 3.6 compares estimates of joint high defaults in the portfolio studied according to two approaches: by assuming normality and by using the best-fit copula. Conclusions are in the last section.

3.2 FINDING THE BEST-FIT COPULA

3.2.1 Parameter estimation techniques

Basically, there are three parametric approaches to estimate copulas from data: the Exact Maximum Likelihood (ML) method[15], the Inference Functions for Margins (IFM), and the Canonical Maximum Likelihood (CML) – see, for instance, Cherubini et al. (2004) and McNeil et al. (2005).

3.2.1.1 The Exact Maximum Likelihood (ML) method

The density (pdf) of the joint distribution F, denoted by f, is given by:

$$f(x_1,..., x_d) = c(F_1(x_1),..., F_d(x_d))\prod_{i=1}^{d} f_i(x_i)$$

where f_i is the univariate density of the marginal distribution F_i and c is the density of the copula given by:

$$c(F_1(x_1),..., F_d(x_d)) = \frac{\partial^d C(F_1(x_1),..., F_d(x_d))}{\partial F(x_1)...\partial F_d(x_d)}$$

[15] Also known as Full Maximum Likelihood (FML).

Assume we have a set of d variables in T time periods represented by $(x_{1t},..., x_{dt})$, $t = 1,...,T$. Let $\mathbf{\Omega} = (\alpha_1,..., \alpha_d, \theta)$ be the parameter vector to be estimated, where $\boldsymbol{\alpha}_i$, $i = 1,..., d$, is the vector of parameters of each marginal distribution F_i and θ is the copula parameter. The log-likelihood function will be:

$$l(\mathbf{\Omega}) = \sum_{t=1}^{T} \ln c(F_1(x_{1t}; \alpha_1),..., F_d(x_{dt}; \alpha_d); \theta) + \sum_{t=1}^{T} \sum_{i=1}^{d} \ln f_i(x_{it}; \boldsymbol{\alpha}_i)$$

The ML estimator $\hat{\mathbf{\Omega}}$ of the parameter $\mathbf{\Omega}$ is found by maximizing the prior function, that is:

$$\hat{\mathbf{\Omega}} = \arg\max l(\mathbf{\Omega})$$

See, for example, Cherubini et al. (2004), McNeil et al. (2005), Charpentier et al. (2007), Genest and Favre (2007), and Trivedi and Zimmer (2005).

3.2.1.2 The method of Inference Functions for Margins (IFM)

In the Exact Maximum Likelihood (ML) approach introduced above, the parameters of the marginals and the copula are estimated together. Joe and Xu (1996) suggested splitting that process into two steps:

(i) estimate parameters $\boldsymbol{\alpha}_i$, $i = 1,..., d$, of the marginal distributions F_i by using the ML method:

$$\hat{\boldsymbol{\alpha}}_i = \arg\max l^i(\boldsymbol{\alpha}_i) = \arg\max \sum_{t=1}^{T} \ln f_i(x_{it}; \boldsymbol{\alpha}_i)$$

where l^i is the log-likelihood function of the marginal distribution F_i;

(ii) estimate a copula parameter θ (or parameter vector $\boldsymbol{\theta}$) by using the results from step (i):

$$\hat{\theta} = \arg\max\, l^c(\theta) = \arg\max \sum_{t=1}^{T} \ln c(F_1(x_{1t};\hat{\alpha}_1),..., F_d(x_{dt};\hat{\alpha}_d);\theta)$$

where l^c is the log-likelihood function of the copula.

For more details, see Cherubini et al. (2004) and Trivedi and Zimmer (2005).

3.2.1.3 The Canonical Maximum Likelihood (CML) method

Like the IFM, this method performs the estimation in two steps. However, it does not make any assumption about the parametric form of the marginal distributions. Following Cherubini et al. (2004), McNeil et al. (2005) and Genest and Favre (2007), the calculations in CML involve the procedures:

(i) transform the dataset $(x_{1t},..., x_{dt})$, t = 1,…,T into uniform variates $(\hat{u}_{1t}..., \hat{u}_{dt})$, using the empirical distributions $\hat{u}_{it} = F_i(x_{it})$ so that we do not need to care about the margins' parameters;

(ii) estimate the copula parameter θ (or parameter vector $\boldsymbol{\theta}$) by maximizing a log-likelihood function that includes the vector $\mathbf{u}_t$ and θ (or $\boldsymbol{\theta}$)**:**

$$\hat{\theta} = \arg\max \sum_{t=1}^{T} \ln c(\hat{u}_{1t},...\hat{u}_{dt};\theta)$$

3.2.1.4 Comparison among parameter estimation techniques

In short, ML (Exact Maximum Likelihood) involves maximizing a function that includes parameters for both the marginal distributions and the copula. The IFM (Inference Functions for Margins) method maximizes two log-likelihood functions. First, the parameters of the margins are found and then these values are used to find the copula parameters. The CML (Canonical Maximum Likelihood) also has two stages but the dataset (default rates in this case) is converted into uniform variables, so that it is not necessary to estimate the margins' parameters. Then, in a second step, the copula parameters are estimated by maximizing a log-likelihood function that includes the uniform variables and the copula parameters.

Although Kole et al. (2007) state that there is no consensus on the best way to fit copulas to data, Durrleman et al. (2000) found the CML to be the best method to model both simulated and real financial data whilst ML and IFM estimations were biased. Genest et al. (2009) state that the IFM approach is less efficient because it is subject to flawed estimations of the univariate distributions (margins) which compromises the second step, namely the search for the copula parameter. Furthermore Cherubini et al. (2004) point out that the ML method tends to be very computationally intensive given that several parameters (for individual distributions and the copula) must be estimated at once.

Therefore this study will employ the CML to estimate the parameters of the copulas candidate to represent the dependence across credit card default rates.

3.2.2 Goodness-of-fit methods

After finding the parameter of each copula, it is necessary to decide which family is the best representation for the data dependence. According to simulations run by Genest et al. (2009) and Berg (2009), the three goodness-of-fit (GoF) methods that presented the highest performances were based on: Empirical Copula (the best performance), Kendall's Transform and Rosenblatt's Transform. These three approaches have become the most popular in copula GoF tests and have been used in other studies (see, for example, Weiss, 2009). All of them check "how far" the data distribution is from the candidate copula distributions and the smaller this "distance" is the more representative the copula is.

3.2.2.1 Test based on the Empirical Copula

Nelsen (2006) explains the empirical copula as a function given by the number of pairs of a dataset that are smaller than or equal to the respective order statistics from the dataset divided by the size of the sample. In his notation:

$$c_n\left(\frac{i}{n},\frac{j}{n}\right)=\frac{\text{number of pairs}\,(\mathrm{x},\mathrm{y})\,\text{in the sample with}\ \mathrm{x}\leq\mathbf{x}_{(i)},\ \mathrm{y}\leq\mathbf{y}_{(j)}}{n}$$

where $\{(\mathbf{x}_k,\mathbf{y}_k)\}_{k=1}^{n}$ denote a sample of size n from a continuous bivariate distribution and $\mathbf{x}_{(i)}$ and $\mathbf{y}_{(j)}$, $1\leq i,j\leq n$, represent the order statistics from the sample.

Let $\mathbf{R}_i$ be the i^{th} rank of a specific variable with n observations ($1\leq i\leq n$). By computing $\mathbf{U}_i=\mathbf{R}_i/(n+1)$, $\mathbf{U}_i$ will be a pseudo-observation equivalent to that rank normalised to (0,1) where the scaling factor (denominator $n+1$) is employed to guarantee $\mathbf{U}_i$ in (0,1). Following these definitions, the empirical copula is given by:

$$\hat{C}(\mathbf{u})=\frac{1}{n+1}\sum_{i=1}^{n}\mathbf{1}(U_{i1}\leq u_1,\ldots,U_{id}\leq u_d)$$

where $\mathbf{u}=(u_1,\ldots u_d)\in[0,1]^d$ is a 1xd row vector that represents the order statistics of the variables, *d* refers to dimensions (the number of variables studied), $\mathbf{U}_i$ is defined as before, $\mathbf{1}$ is an indicator function that returns 1 if all conditions in parenthesis are satisfied and 0 otherwise, and n is the number of observations.

This goodness-of-fit procedure aims to test the null hypothesis that the underlying copula *C* belongs to the family of a candidate copula C_0, i.e.,

$H_0^E : C \in C_0$ where the superscript E is used to indicate that the null hypothesis refers to the Empirical Copula GoF method. The procedure for a copula family $C_{\hat{\theta}}$ with estimated parameter $\hat{\theta}$ involves the calculation of the empirical process $E = \sqrt{n}(\hat{C} - C_{\hat{\theta}})$ for n observations. Applying the Cramér-von Mises statistic (Genest et al., 2009 and Berg, 2009), the test becomes:

$$\hat{E} = n\int_{[0,1]^d} \{\hat{C}(\mathbf{u}) - C_{\hat{\theta}}(\mathbf{u})\}^2 d\hat{C}(\mathbf{u}) = \sum_{i=1}^{n} \{\hat{C}(\mathbf{u}_i) - C_{\hat{\theta}}(\mathbf{u}_i)\}^2$$

where $\hat{C}$ is the empirical copula defined above and $C_{\hat{\theta}}$ is the candidate copula. Since this test evaluates the “distance” between the dataset distribution (here using the empirical copula as a proxy) and the estimated copula, larger values of $\hat{E}$ entail higher chances of rejection of the tested copula.

Genest and Rémillard (2008) demonstrate that goodness-of-fit procedures based on this approach are asymptotically consistent. Then when $C \notin C_0$, the probability of H_0 being rejected goes to 1 when $n \to \infty$.

Genest et al. (2009) explain that the limiting distribution of $\hat{E}$ depends on the unknown copula and on its parameter θ. Hence the asymptotic distribution of $\hat{E}$ cannot be calculated and p-values must be inferred by means of bootstrap methods.

3.2.2.2 Test based on Kendall’s Transform

This approach uses a transformation of Kendall’s multivariate coefficient of concordance. The formulae presented below are based on Genest et al. (2009) and Berg (2009). The first step is to calculate Kendall’s dependence function:

$$K_T(w) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{C(U_i,..., U_d) \leq w\} = \Pr[C(U_i,..., U_d) \leq w]$$

where $w = i/(n+1)$, $w \in [0,1]$, and the other notations follow the definitions above.

By replacing the copula in the function above with the empirical copula (defined before) we get an empirical version of Kendall's function:

$$\hat{K}_T(w) = \frac{1}{n+1}\sum_{i=1}^{n} \mathbf{1}\{\hat{C}(u_i,..., u_d) \le w\} \cong \Pr[\hat{C}(u_i,..., u_d) \le w]$$

The null hypothesis in this test is different from the one in the previous method. Here, the null hypothesis is that Kendall's Transform of the unobserved copula (K) belongs to a set of Kendall's Transforms (K_0) related to the candidate copula, that is, $H_0^K : K \in K_0$ where the superscript K denotes the Kendall's Transform GoF method. Under this supposition, $K_T = \hat{K}_T$.

The test follows the empirical process $K = \sqrt{n}(K_T - \hat{K}_T)$ that refers to the difference between Kendall's Transform of the candidate copula and the empirical copula, K_T and $\hat{K}_T$ respectively.

Applying the previous notation, the Cramér-von Mises statistic for this approach is:

$$\hat{K} = n\int_0^1 \{K_T(w) - \hat{K}_T(w)\}^2 d\hat{K}_T = \sum_{i=1}^{n}\left\{K_T\left(\frac{i}{n+1}\right) - \hat{K}_T\left(\frac{i}{n+1}\right)\right\}^2$$

It is important to note that Genest et al. (2009) compare the null hypotheses of the empirical copula and the Kendall's Transform and highlight that $H_0^E \subset H_0^K$. That is, the nonrejection of H_0^K may not imply the nonrejection of H_0^E. As a result, in general, tests based on this Kendall approach are not (asymptotically) consistent.

Those authors also point out that the asymptotic distribution of $\hat{K}$ depends on the underlying copula and on its parameter; therefore p-values related to that statistic are found through simulations.

3.2.2.3 Test based on Rosenblatt's Transform

The Rosenblatt's Transform of a copula maps a vector $\mathbf{u} = (u_1, ...u_d) \in (0,1)^d$ to a function $\Re(\mathbf{u}) = (e_i, \ldots, e_d) \in (0,1)^d$ where $e_1 = u_1$. For each dimension $i \in \{2,..., d\}$, e_i is defined as (see Genest et al., 2009):

$$e_i = \frac{\partial^{i-1} C(u_1,...,u_i,1,...,1)}{\partial u_1 ... \partial u_{i-1}} \Bigg/ \frac{\partial^{i-1} C(u_1,...,u_{i-1},1,...,1)}{\partial u_1 ... \partial u_{i-1}}$$

The test derived from this transformation comes from the property that a vector $\mathbf{U}$ is distributed as a copula C (i.e. $\mathbf{U} \sim$ C) if and only if the distribution of $\Re(\mathbf{U})$ is an independent copula $C_{\perp} = (e_i, \ldots, e_d) = e_i * \ldots * e_d$, with $e_i, \ldots, e_d \in [0,1]$.

So the null hypothesis $H_0^R : \mathbf{U} \sim C \in C_0$ corresponds to $H_0^R : \Re_\theta(\mathbf{U}) \sim C_\perp$ according to a parameter θ estimated for the candidate copula[16].

To test this hypothesis, we use the fact that H_0^R entails the interpretation of the pseudo-observations $\mathbf{E}_d = \Re_\theta(\mathbf{U}_d)$ as a sample from $C_\perp$, the independent copula.

An empirical distribution function related to this process is:

$$D_n(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^{n} 1(E_i \le u), \quad u \in [0,1]^d$$

[16] The superscript R is used to indicate that this null hypothesis pertains to the test based on Rosenblatt's Transform.

which, under H_0^R, is expected to be "close" to C$_\perp$.

The test is based on one of the Cramér-von Mises statistics[17] of the process $R = \sqrt{n}(D_n - C_\perp)$:

$$\hat{R} = n\int_{[0,1]^d} \{D_n(\mathbf{u}) - C_\perp(\mathbf{u})\}^2 du = \frac{n}{3^d} - \frac{1}{2^{d-1}} \sum_{i=1}^{n} \prod_{k=1}^{d} (1 - E_{ik}^2) + \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \prod_{k=1}^{d} (1 - (E_{ik} \vee E_{jk}))$$

where $E_{ik} \vee E_{jk} = \max(E_{ik}, E_{jk})$.

As in the two approaches described above, the asymptotic distribution of this statistic is conditional on the latent copula and its parameter. Consequently, p-values for the test are generated from bootstrap techniques.

3.2.2.4 Test procedures

The statistic for each test, $\hat{E}$, $\hat{K}$, and $\hat{R}$ will be calculated and the smallest value will indicate the best-fit according to the respective method. Thus, for each pair of segments from the credit card portfolio analysed, three copulas will be designated as potential best-fit representations of its dependence structure. It is possible (and expected) that the three approaches yield conflicting results in some cases (i.e. two or three distinct copulas for the same pair). In these circumstances, the Empirical-based test must be considered more reliable since, according to Genest et al. (2009), it is the method that presents the least data transformation and its superiority was confirmed by Berg (2009).

In order to verify the significance of the GoF tests, Genest et al. (2009) and Berg (2009) present some routines to calculate p-values concerning the null hypothesis that the dataset dependence (underlying copula) is equal to the tested copula.

[17] Genest et al. (2009) present two Cramér-von Mises statistics for the Rosenblatt's approach. The statistic used in this thesis is the one that had higher performance in the tests run by those authors.

The procedure to find the p-values consists of simulating the candidate copulas many times with their respective parameters found (via CML – as explained in Section 3.2.1) and checking which proportion of them is “farther” from the empirical data than the candidate copula with the exact parameter found via maximisation. Thus high p-values suggest that the considered copula cannot be rejected because it is closer to the observed dataset than most of the other simulated copulas.

3.3 DATA DESCRIPTION

This empirical study is based on a random sample from the credit card portfolio of a large UK bank comprising the monthly payment status of 177,234 accounts over the period April/2007 – March/2009. The dataset initially was composed of 350,066 credit card loans but the credit scores for 77,780 accounts were not available and another 95,052 accounts were open after the first month covered by the dataset. This means that the default rates over the two years studied were calculated based only on the loans existing in April/2007 with an available credit score.

The dataset was split into five segments according to the loans’ credit quality (credit score provided by the Bank) in the first month (April/2007). Each segment corresponds to a quintile of the score distribution such that the least risky segment (named “A”) presents the highest scores and the risk level increases with the reduction of the scores up to the riskiest segment (called “E”).

Some of the accounts were dormant (310 in segment A, 269 in segment B, 39 in segment C, and 0 in segments D and E) but they represented only 0.35% of the credit card loans effectively analysed and therefore these dormant accounts did not compromise the calculations of the default rates and our subsequent conclusions.

Default was defined as the non-payment of three monthly instalments (consecutively or not) conditional on no prior default. Thus the default rate for each segment in a particular month was calculated as the amount of loans that reached their third month in arrears for the first time (i.e. conditional on no prior default) divided by the number of active accounts in that month. Once a loan defaulted it was excluded from the dataset (i.e. default is considered an absorbing state).

This procedure generated a time series of default rates with 24 observations for each segment and these values were used to estimate the dependence between the segments.

The summary statistics of the data is shown in Table 3.1. For confidentiality reasons, the monthly default rates of each segment and the number of accounts defaulted in each month are not reported.

Table 3.1: Summary statistics of default rates for the five segments of the credit card portfolio

SEGMENT	MEAN	STANDARD DEVIATION	SKEWNESS	KURTOSIS
A	0.00015	0.00009	-0.22504	2.55931
B	0.00056	0.00035	-0.04654	2.17929
C	0.00295	0.00141	-0.57593	2.75557
D	0.01117	0.00270	-2.78037	12.64415
E	0.03226	0.01873	2.12537	8.45527

Data refers to April/2007 – March/2009.

As expected, the mean default rates increase with decreases in the credit quality (from segment A to E). The data dispersion (measured by the standard deviation) has similar behaviour.

According to the two rightmost columns, the three first segments are closer to the normal distribution than the two riskiest ones (A, B and C have skewness and kurtosis relatively closer to zero and three, respectively, when compared to segments D and E). This is confirmed by the Jarque-Bera test which tests the null hypothesis that a sample comes from a normal distribution (see Table 3.2 where higher values of the Jarque-Bera statistics lead to the rejection of the null hypothesis).

Table 3.2: Jarque-Bera* test for the default rates' segments

SEGMENT	CAN VALUES BE APPROXIMATED TO NORMAL?	JARQUE-BERA STATISTICS
A	Yes	0.39678
B	Yes	0.68222
C	Yes	1.38653
D	No	123.93139**
E	No	47.82875**

* This test checks if the normal distribution is a good approximation for the data analysed.
** These results are significant at the 1% level (p-value < 0.001 in both cases).

The fact that three of the segments (A, B and C) may be satisfactorily represented by the normal distribution (albeit this result is not significant) does not imply that the dependence between pairs involving two of those segments will be better expressed by the Gaussian (Normal) Copula than by another copula. As it will be shown ahead, even normal data may have a diverse dependence structure.

As said before, 95,052 accounts open after the first month (April/2007) were eliminated from the dataset. This was done because the inclusion of those accounts would distort our definition of segments (namely, equally-sized cohorts of loans concerning the percentiles of the credit score). If the accounts open after the first month were inserted into the analysis and the credit score bounds for each segment were kept constant, the number of distinct scores would be likely different in the five segments. So, the concept of segments would be changed inasmuch as each of them would cover an unbalanced number of percentiles of the data. Moreover the number of new accounts in a specific segment could be much lower (higher) than the number of new loans in the other groups such that its default rates would be artificially modified. This would not be due to the increase (reduction) of the payment failures but instead be a consequence of the size of the segment (specially in the first two months after the new accounts were added to the portfolio because, according to our criterion, they would not default before their third month).

On the other hand, if the new accounts were incorporated into the dataset and the segments were kept in the consecutive ranges of 20 percentiles, the limiting values of the credit scores for each segment would possibly change and some loans with a stable score (i.e. consumers with invariable characteristics assessed by the Bank[18]) would move across segments. In this fashion, the risk (score level) of the segments would not be constant and the dependence structures estimated would represent connections between different consumers in each month. Thus we would not know whether the oscillations in the dependence strength were due to the association between pairs of segments or were an effect of the distinct composition of the segments every month.

Considering that the default rates are given by the number of defaulted accounts in the respective month divided by the total of active accounts in that month, it is clear that the non-inclusion of new accounts leads to a bias (as time goes on) in the calculation of default rates. For a particular group, the same number of defaults in two different months will result in a higher default rate in the later month (since it will have fewer active accounts). The worst effect of this drawback on the empirical analysis could happen if few occurrences of default in a final month with reduced accounts resulted in higher default rates than many occurrences in an initial month with more loans in the portfolio. This would affect the ranks (percentiles) of the default rates and, consequently, would influence the best-fit copula estimated from the data. However, as implied by the mean default rates in Table 3.1, the number of defaults is considerably smaller than the number of active accounts. Therefore the reduction of the denominator (total of active accounts) in the ratio that

[18] The variables considered by the Bank were not informed but may include, for example, income, outstanding debt, employment condition and housing status.

represents the default rates is minimal over time[19] and this guarantees that the order of the default rates is almost surely not changed which implies that the copulas estimated are not influenced and the bias is negligible in this case.

3.4 FAMILIES OF CANDIDATE COPULAS

Ten copula families are tested to represent the dependence across default rates. They were selected from the bivariate one-parameter families described in Joe (1996, Chapter 5)[20] along with the Student t Copula.

Table 3.3 lists them and their main features in terms of structure. The cumulative distribution functions (*cdfs*) of the candidate copulas were presented in Chapter 2, Section 2.3.2.

Table 3.3: Candidate copulas and their respective features

COPULA	DEPENDENCE STRUCTURE
Gaussian	Symmetric dependence without tail dependence
Frank	Symmetric dependence without tail dependence
FGM(*)	Symmetric dependence without tail dependence
Placket	Symmetric dependence without tail dependence
Student t	Symmetric dependence with tail dependence
Clayton	Left (lower) tail dependence
Gumbel	Right (upper) tail dependence
Galambos	Right (upper) tail dependence
Hüsler-Reiss	Right (upper) tail dependence
Joe	Right (upper) tail dependence

(*) FGM stands for Farlie-Gumbel-Morgenstern.

In sum, the first four copulas in Table 3.3 indicate that the variables (default rates, in this study) have the same level of dependence below and above their mean and there is no higher association (when compared to the multivariate

[19] That is, the active accounts minus the defaults in the preceding month are relatively close to the initial number of accounts. Since the default *rates* tend to follow the number of defaults in each month, the ranks (percentiles) of both will be the same.

[20] Only the absolutely continuous families were chosen (nine families). This property is desirable in order to simulate such copulas (used in the goodness-of-fit tests; see, for example, Joe, 1996, p. 146) and estimate their parameters. Both procedures demand derivations of the copulas' *cdfs*.

normal distribution) among extreme values. The Student t is also symmetric but points out a more intense relationship among extreme events (compared to the first four families). The Clayton copula indicates that smaller values are more linked and the other four copulas express the opposite: higher values are more associated. This last case is the one that brings more concern with respect to losses in credit portfolios since it means that higher default rates tend to happen together more often, i.e. higher losses in each segment occur at the same time and the lender is more subject to financial deficits.

3.5 DEPENDENCE STRUCTURE IN THE CREDIT CARD PORTFOLIO

3.5.1 Estimation based on the complete default rate distributions

To find the best-fit copulas for the ten pairs of segments, the parameters of the candidate copulas were estimated for each pair according to the Canonical Maximum Likelihood method which, according to the literature, outperforms the other approaches mentioned in Section 3.2.1. Then the three goodness-of-fit (GoF) approaches presented in Section 3.2.2 were used to define which copula better represents the dependence in each pair.

Table 3.4 displays the best copulas (in the upper-right triangle) along with the linear correlations (in the lower-left triangle). The copulas displayed for each pair were estimated according to the Empirical Method (first family shown), Kendall’s Transform (in parenthesis), and Rosenblatt’s Transform (in square brackets). The parameters of the best-fit copulas are reported in Appendix A.

The *rejection* level of the estimates is indicated by ** and * which represent the levels 5% and 10%, respectively (for instance, although the hypothesis of Clayton Copula for the pair AB based on the Empirical GoF approach is the best among the ten alternatives, it can be rejected at the 5% level, i.e. with 95% of confidence). Appendix B contains tables with the outcome of all three GoF approaches for estimations that used the whole default distributions and includes their respective p-values.

Table 3.4: Best-fit copulas based on the complete distributions

SEGMENTS	A	B	C	D	E
A	1	Clayton (Plackett) [Frank]	Clayton (Plackett) [Clayton]	Galambos (Galambos) [Gumbel**]	Student t (Frank) [Plackett]
B	0.7375	1	Clayton (Clayton) [Plackett]	Hüsler-Reiss (Galambos) [Joe*]	Student t (Plackett*) [Plackett]
C	0.7888	0.9536	1	Hüsler-Reiss (Joe) [Joe*]	Gaussian (Frank**) [Plackett]
D	0.3730	0.4598	0.5653	1	Plackett (FGM*) [Plackett]
E	-0.4966	-0.4916	-0.5217	0.1241	1

Best-fit copulas (upper-right triangle) and linear correlation (lower-left triangle) for default rates of pairs of segments (estimation based on the best fit to complete distributions). The copulas displayed for each pair of segments are respectively based on Empirical Copula, Kendall's Transform (in parenthesis), and Rosenblatt's Transform (in square brackets).
* and ** indicate copulas with the highest probability of nonrejection, i.e. results with the highest significance (p-values between 0.90 and 0.95 and p-values greater than 0.95, respectively).

The analysis of the results will be based on the copulas estimated following the Empirical Copula method given that it was found in the literature to be the most robust amongst the three models considered in this study (see, as mentioned in Section 3.2.2, Berg, 2009 and Genest et al., 2009).

Eight pairs have tail dependence albeit in two cases the estimations can be rejected at the 5% and 10% levels (pairs AB and AC, respectively; see p-values in Tables B.1 and B.2, Appendix B). Thus, in most of the cases, the link across extreme default rates (lower and/or higher) is stronger than assumed by the Gaussian Copula (which is implicit in traditional credit risk models).

Three of the pairs (AD, BD, and CD) exhibit right-tail dependence, meaning that higher default rates are more associated than the other levels which may strengthen the Bank's losses in downturns. Three other pairs (AB, AC, and BC) have more intense relationships among low default rates, expressing the most profitable scenario for the Bank inasmuch as most of its debtors tend to keep their repayments simultaneously in upturns whilst delinquencies are not very related in downturns.

Pairs AE and BE may present those two effects on the Bank's results. The symmetric tail association represented by the Student t Copula implies that both lower and higher ranks of defaults are more associated than intermediate rates.

The two riskiest segments (CE and DE) are not tail dependent. This condition is beneficial for the lender because the highest default levels do not get more linked in downturns.

In contrast to all other pairs, the pairs involving segment E have symmetric dependence structures. This suggests that the association of that riskiest segment with the other loans has similar intensity in opposite economic scenarios (booms or crashes). Thus the same level of connections among the riskiest debtors and the other loans in downturns (that raise the losses) may be expected in upturns (so that losses are reduced in those periods and profits are potentially amplified).

It is interesting to note that even when the individual distributions (the default distribution for each segment in this study) are satisfactorily approximated by the normal distribution (see Table 3.2), their joint distribution may not be expressed by the Gaussian Copula. This is the case of the pairs AB, AC, and BC (although the best-fit copula of the first two pairs can be rejected).

The p-values for the goodness-of-fit tests displayed in Appendix B entail the non-rejection of many copula families which is possibly a consequence of the short dataset used (24 observations) that is not enough to form unambiguous patterns that match unique joint distributions for each pair of segments. Thus the main inference from this empirical analysis is not the rejection of the Gaussian dependence but the identification of other families that may represent the dependence among credit card loans more accurately and improve the estimation of joint extreme defaults. The dependence in nine of the pairs is better expressed by other copulas; only the pair CE has the Gaussian as the best-fit copula.

3.5.2 Estimation based on the upper tails of default rate distributions

Since the main purpose of estimating the dependence structure is to calculate joint “high” defaults, it is possible that a better performance of the copula approach may be achieved if the best-fit copulas are found considering only the right tail of the default rate distributions. In this section, we estimate the copula whose right tail (here, defined as above the 75th percentile of each marginal variable u_1 and u_2 – following the notation in Section 3.2.2.1) gives the best fit to the right tail of the empirical distribution. In this bivariate case, the Empirical Copula $\hat{C}_R$ limited to “high” percentiles (above 0.75 in our example) is calculated from the dataset as:

$$\hat{C}_R(\mathbf{u} \mid \mathbf{u} \geq 0.75) = \frac{1}{n+1}\sum_{i=1}^{n} \mathbf{1}(U_{i1} \leq u_1, U_{i2} \leq u_2)$$

where the term $\mathbf{u} \geq 0.75$ indicates that both marginals u_1 and u_2 must be equal to or greater than 0.75. We recognise that this percentile is too low for typical definitions of “high” events but it was chosen because we needed to specify a certain number of extreme occurrences in each segment that allowed us to observe joint events across high defaults and even though we defined only the six highest values of each default distribution as its extreme occurrences, this number represents 25% of each segment’s distribution.

Each candidate copula, $C_{R\hat{\theta}}$, pertaining to this test applied to the rightmost region of the default distributions will be evaluated for the same percentiles considered in the Empirical Copula $\hat{C}_R$ and will be given by[21]:

$$C_{R\hat{\theta}} = C_{\hat{\theta}}(\mathbf{u} \mid \mathbf{u} \geq 0.75)$$

where $C_{\hat{\theta}}$ is a candidate copula with a parameter $\hat{\theta}$ estimated according to the procedures described in Section 3.2.1 and limited, in this example, to percentiles equal to or higher than 0.75.

Then we use C_R and $C_{R\hat{\theta}}$ to calculate the Cramér-von Mises statistic, $\hat{E}_R$ (as in Section 3.2.2.1):

$$\hat{E}_R = n\int_{[0,1]^d} \{\hat{C}_R(\mathbf{u}) - C_{R\hat{\theta}}(\mathbf{u})\}^2 d\hat{C}_R(\mathbf{u}) = \sum_{i=1}^{n} \{\hat{C}_R(\mathbf{u}_i) - C_{R\hat{\theta}}(\mathbf{u}_i)\}^2$$

where $\mathbf{u} \geq 0.75$. As a consequence, this test will point out the copula family that has the area in the right tail (when the percentiles u_1 and/or u_2 are greater than or equal to 0.75) closest to the area in the right tail of the joint distribution of the observed data. The other tests (Kendall's Transform and Rosenblatt's Transform) followed the same idea as described for C_R, $C_{R\hat{\theta}}$ and $\hat{E}_R$.

Keep in mind that that the area not considered in these goodness-of-fit tests (75% of the data in the left side of the distributions) has no impact on the dependence estimated for the right tails and therefore this right-tail dependence may have any shape irrespective of the dependence of the lowest 75% percentiles.

[21] Note that the "tail" of the distributions we are modeling is formed not only by the concurrent events $u_1 \geq 0.75$ and $u_2 \geq 0.75$ but also by $u_1 \geq 0.75$ and $u_2 < 0.75$ (and vice versa).

The best-fit copulas selected in accordance with GoF tests based on complete default distributions (as in Section 3.5.1) are supposed to be the best representation of joint distributions in general (i.e. for all values of u_1 and u_2) and may not be the best approximation of the upper tail specifically (which can be given by another copula family). On the other hand, copulas chosen according to $\hat{E}_R$ are the best approximation of "high" values of default rates and might not be the best representation of default rates smaller than the respective 75th percentiles. This approach seems to be an original way to estimate copulas to express joint high events since such strategy has not been found in the literature.

The best-fit copulas based on this alternative method are displayed in Table 3.5 and their respective parameters are in Appendix C.

Table 3.5: Best-fit copulas based on the best fit to the right-hand tails

SEGMENTS	A	B	C	D	E
A	1	Galambos (Galambos) [Frank]	Joe* (Galambos) [Clayton]	Galambos (Gumbel) [Gumbel*]	Frank (Clayton) [Plackett]
B	0.7375	1	Clayton (Student t) [Plackett]	Gumbel (Clayton) [Joe*]	Frank (Clayton) [Plackett]
C	0.7888	0.9536	1	Joe (Clayton) [Joe]	Plackett** (Clayton) [Plackett]
D	0.3730	0.4598	0.5653	1	Joe** (Joe) [Plackett]
E	-0.4966	-0.4916	-0.5217	0.1241	1

Best-fit copulas (upper-right triangle) and linear correlation (lower-left triangle for default rates of pairs of segments (correlation for complete default distributions and copula estimation based on the best fit to right tails. The copulas displayed for each pair of segments are respectively based on Empirical Copula, Kendall's Transform (in parenthesis), and Rosenblatt's Transform (in square brackets).
* and ** indicate copulas with the highest probability of nonrejection, i.e. results with the highest significance (p-values between 0.90 and 0.95 and p-values greater than 0.95, respectively).

Similar to the previous table, the copulas for each pair of segments are assessed from the Empirical Method (first family displayed for each pair), Kendall's Transform (in parenthesis), and Rosenblatt's Transform (in square brackets). The results for the three GoF approaches are detailed in Appendix D which also reports the p-values calculated.

The results based on the Empirical Copula Method[22] reveal that seven out of the ten pairs present tail dependence (AB, AC, AD, BC, BD, CD, DE), i.e. the association between default rates in extreme cases is more intense than assumed by the Gaussian Copula. Six of the pairs have right-tail dependence (AB, AC, AD, BD, CD, DE) and only one pair, CD, has low default rates more related. Therefore these estimations indicate that most of the associations among credit card loans lead to accentuated losses in adverse scenarios.

The pairs that did not present tail dependence were exactly the ones with negative dependence (all of them involving the riskiest segment, E, which is advantageous for the Bank since its highest expected losses are mitigated by better performance of other segments). Note that even in these three instances the best-fit copula was not the Gaussian one. This has some implications only in the central region of the default distribution and does not impact investigations concentrated in extreme events (which is the case of this study).

Compared to the estimations in the prior section (founded on the complete default distributions), the results derived from the fit to the right tail allowed a greater rejection of the Gaussian copula (especially with reference to the Kendall's Transform GoF approach according to which that copula can be rejected in two pairs when using the complete default distributions and in six pairs when only the right tails of those distributions are considered).

[22] As before, the estimations are interpreted with respect to this approach due to its superior robustness according to the pertinent literature (Genest et al., 2009 and Berg, 2009).

3.6 ESTIMATION OF JOINT EXTREME DEFAULTS: COMPARISON BETWEEN TRADITIONAL METHODS AND COPULAS

In this section, estimations of joint losses (default rates) following the assumption of normality (both univariate and multivariate) are compared to evaluations based on Copula Theory. It is expected that, in general, approaches based on copulas give more accurate assessment of joint extremely high losses. To test this hypothesis, we calculated the probability of default rates, X_I and X_J in segments I and J respectively, being simultaneously above specific levels (values) x_I and x_J as follows:

- Assuming normality:

$$\Pr(X_I > x_I, X_J > x_J) = 1 - \Phi(x_I) - \Phi(x_J) + \Phi(x_I, x_J)$$

where Φ indicates the *cdf* of a normal distribution; and

- Using the best-fit survival copula:

$$\Pr(X_I > x_I, X_J > x_J) = \hat{C}[1 - F_I(x_I), 1 - F_J(x_J)]$$

where $\hat{C}$ is a survival copula, i.e. links "survival ranks": $1 - F(.)$; F_I and F_J are the *cdfs* of the (unknown) distributions of default rates X_I and X_J, respectively.

Given that the dataset has 24 observations, the following proportions of "extremely high" levels (percentiles) of default rates were selected: 4.17% (1/24), 8.33% (2/24), 12.50% (3/24), 16.67% (4/24), 20.83% (5/24), and 25% (6/24). Thus, for each pair of segments we compare estimations of potential joint losses in those highest levels. For instance, the likelihood that the 4.17% highest default rates in segment I happen at the same time that the 4.17% highest default rates in segment J, and so on.

As the best-fit copulas were estimated according to two approaches (based on the complete default distributions and on their right tails), the survival copulas used to evaluate the joint occurrences were also determined following both strategies.

3.6.1 Survival copulas estimated considering the complete distributions of default rates

The results pertaining to the survival copulas estimated from the whole default distributions are presented in Table 3.6 where the first column shows the *proportion* of the highest defaults (not the default rates themselves). The columns labelled “Dataset” give the proportion of joint default rates observed in the credit card portfolio at the respective levels.

According to Table 3.6, in 63.33% of the scenarios, the approximations derived from copulas were closer to the likelihood of simultaneous high default rates observed in the credit card portfolio than the predictions from the normal distribution. That is, in 38 out of the 60 situations represented in Table 3.6, the absolute difference between columns “Dataset” and “Copula” was smaller than the absolute difference between columns “Dataset” and “Normal”.

Notwithstanding, this approach resulted in higher underestimation rate (16.67%) than calculations assuming normality (11.67%). Hence this empirical analysis does not support the hypothesis that evaluations based on normality assumptions are prone to underestimate joint “extreme” defaults since their underestimation ratio was relatively low (11.67%) compared to the alternative method (16.67%). This is likely due to the short period covered by the dataset which virtually ruled out the probability of *joint* “extreme” occurrences (if we defined “extreme” in this dataset as, for example, above the 95th percentile, such “extreme” events would take place only if the highest default rate of each segment happened in the same month).

Table 3.6: Comparison of predicted joint extreme default using entire samples

Panel A: pair AB (Joe)

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00996	0.00579	0.00996	0.00579
8.33%	0.00000	0.03390	0.02110	0.03390	0.02110
12.50%	0.04167	0.05466	0.04363	0.01299	0.00196
16.67%	0.04167	0.08998	0.07180	0.04831	0.03013
20.83%	0.04167	0.09485	0.10445	0.05318	0.06278
25.00%	0.12500	0.11761	0.14070	0.00739	0.01570
Total difference for this pair				0.16573	0.13747

Panel B: pair AC (Gumbel)

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.01429	0.01219	0.01429	0.01219
8.33%	0.00000	0.04202	0.03188	0.04202	0.03188
12.50%	0.04167	0.06583	0.05593	0.02416	0.01427
16.67%	0.04167	0.10950	0.08335	0.06783	0.04169
20.83%	0.08333	0.12797	0.11358	0.04464	0.03025
25.00%	0.12500	0.15524	0.14625	0.03024	0.02125
Total difference for this pair				0.22320	0.15152

Panel C: pair AD (Galambos)

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00513	0.00174	0.00513	0.00174
8.33%	0.00000	0.03551	0.00694	0.03551	0.00694
12.50%	0.00000	0.04552	0.01563	0.04552	0.01563
16.67%	0.00000	0.09269	0.02778	0.09269	0.02778
20.83%	0.04167	0.10413	0.04340	0.06247	0.00174
25.00%	0.04167	0.14683	0.06250	0.10516	0.02083
Total difference for this pair				0.34647	0.07465

Panel D: pair AE (Student t)

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00000	0.00074	0.00000	0.00074
8.33%	0.00000	0.00006	0.00247	0.00006	0.00247
12.50%	0.00000	0.00307	0.00547	0.00307	0.00547
16.67%	0.00000	0.01481	0.01017	0.01481	0.01017
20.83%	0.00000	0.02134	0.01706	0.02134	0.01706
25.00%	0.00000	0.05369	0.02672	0.05369	0.02672
Total difference for this pair				0.09297	0.06263

Panel E: pair BC (Joe)

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.04167	0.02615	0.01446	0.01552	0.02720
8.33%	0.08333	0.05749	0.04447	0.02584	0.03887
12.50%	0.08333	0.10913	0.08128	0.02580	0.00206
16.67%	0.08333	0.11965	0.12152	0.03632	0.03818
20.83%	0.16667	0.12932	0.16359	0.03734	0.00308
25.00%	0.16667	0.14136	0.20665	0.02531	0.03998
Total difference for this pair				0.16612	0.14937

Panel F: pair BD (Student t)

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00783	0.00720	0.00783	0.00720
8.33%	0.04167	0.03707	0.01838	0.00460	0.02328
12.50%	0.04167	0.06548	0.03276	0.02382	0.00891
16.67%	0.04167	0.07776	0.05012	0.03609	0.00845
20.83%	0.04167	0.08322	0.07039	0.04155	0.02872
25.00%	0.04167	0.09841	0.09352	0.05674	0.05186
Total difference for this pair				0.17062	0.12842

Panel G: pair BE (Student t)

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00000	0.00113	0.00000	0.00113
8.33%	0.00000	0.00005	0.00307	0.00005	0.00307
12.50%	0.00000	0.00503	0.00601	0.00503	0.00601
16.67%	0.00000	0.00873	0.01026	0.00873	0.01026
20.83%	0.00000	0.01177	0.01627	0.01177	0.01627
25.00%	0.00000	0.02265	0.02459	0.02265	0.02459
Total difference for this pair				0.04822	0.06133

Panel H: pair CD (Clayton)

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.01400	0.00599	0.01400	0.00599
8.33%	0.04167	0.05045	0.01583	0.00878	0.02584
12.50%	0.04167	0.08756	0.02871	0.04590	0.01296
16.67%	0.04167	0.10206	0.04444	0.06039	0.00278
20.83%	0.04167	0.12320	0.06296	0.08153	0.02129
25.00%	0.08333	0.13932	0.08423	0.05598	0.00090
Total difference for this pair				0.26658	0.06975

Panel I: pair CE (Gaussian)

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00000	0.00006	0.00000	0.00006
8.33%	0.00000	0.00005	0.00060	0.00005	0.00060
12.50%	0.00000	0.00577	0.00231	0.00577	0.00231
16.67%	0.00000	0.01001	0.00590	0.01001	0.00590
20.83%	0.00000	0.01761	0.01209	0.01761	0.01209
25.00%	0.04167	0.03167	0.02155	0.00999	0.02011
Total difference for this pair				0.04342	0.04107

Panel J: pair DE (Plackett)

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00002	0.00517	0.00002	0.00517
8.33%	0.00000	0.00923	0.01797	0.00923	0.01797
12.50%	0.00000	0.06704	0.03605	0.06704	0.03605
16.67%	0.00000	0.09455	0.05816	0.09455	0.05816
20.83%	0.04167	0.11596	0.08357	0.07429	0.04190
25.00%	0.12500	0.16891	0.11179	0.04391	0.01321
Total difference for this pair				0.28904	0.17246

Comparison between estimations of likelihood of joint extremely high default rates (normality vs. copulas). The survival copulas are informed in parenthesis after the names of the pairs and were estimated based on entire distributions of default rates.

However, if we take into account the average magnitude of the differences between estimations and observed default rates *in each pair* of segments, the copula method was better for nine of the ten pairs (the sum of the fifth column is greater than the sum of the last column for all pairs apart from BE). In other words, “on average” (considering the six risk levels tested), the copula model yielded better results.

3.6.2 Survival copulas estimated considering the right tail of default rate distributions

The comparison between copula and traditional methods was repeated by using survival copulas estimations based only on the upper tail (above the 75th percentile) of the default rate distributions (method similar to the one presented in Section 3.5.2). The results are displayed in Table 3.7.

The copula estimations were closer to the real default rates (observed in the dataset) in 70% of the cases but presented a higher underestimation rate (26.67%) than the normality-based estimations (11.67%). As in the analysis of the previous item (for survival copulas estimated from the whole default distributions), this finding does not corroborate the idea that evaluations from normality assumptions tend to underestimate the odds of extreme events. Again, this failure in confirming that hypothesis is likely due to the short range covered by the dataset which excludes the possibility of checking simultaneous occurrences in the very tail of the distributions (for instance, the 1% highest default rates).

Nevertheless, taking into consideration each pair separately, the average of the difference between estimations and observed default rates was smaller for the copula approach in all ten pairs (for each pair, the sum of the last column is less than the sum of the fifth column in Table 3.7).

Table 3.7: Comparison of predicted joint extreme default rates using tail distributions

Panel A: pair AB (Clayton)

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00996	0.01015	0.00996	0.01015
8.33%	0.00000	0.03390	0.02350	0.03390	0.02350
12.50%	0.04167	0.05466	0.03934	0.01299	0.00232
16.67%	0.04167	0.08998	0.05753	0.04831	0.01586
20.83%	0.04167	0.09485	0.07800	0.05318	0.03633
25.00%	0.12500	0.11761	0.10076	0.00739	0.02424
Total difference for this pair				0.28904	0.13988

Panel B: pair AC (Clayton)

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.01429	0.01101	0.01429	0.01101
8.33%	0.00000	0.04202	0.02504	0.04202	0.02504
12.50%	0.04167	0.06583	0.04145	0.02416	0.00021
16.67%	0.04167	0.10950	0.06010	0.06783	0.01844
20.83%	0.08333	0.12797	0.08095	0.04464	0.00239
25.00%	0.12500	0.15524	0.10399	0.03024	0.02101
Total difference for this pair				0.22320	0.07809

Panel C: pair AD (Galambos)

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00513	0.00174	0.00513	0.00174
8.33%	0.00000	0.03551	0.00694	0.03551	0.00694
12.50%	0.00000	0.04552	0.01563	0.04552	0.01563
16.67%	0.00000	0.09269	0.02778	0.09269	0.02778
20.83%	0.04167	0.10413	0.04340	0.06247	0.00174
25.00%	0.04167	0.14683	0.06250	0.10516	0.02083
Total difference for this pair				0.34647	0.07465

Panel D: pair AE (Frank)

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00000	0.00024	0.00000	0.00024
8.33%	0.00000	0.00006	0.00112	0.00006	0.00112
12.50%	0.00000	0.00307	0.00292	0.00307	0.00292
16.67%	0.00000	0.01481	0.00600	0.01481	0.00600
20.83%	0.00000	0.02134	0.01085	0.02134	0.01085
25.00%	0.00000	0.05369	0.01806	0.05369	0.01806
Total difference for this pair				0.09297	0.03919

Panel E: pair BC (Joe)

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.04167	0.02615	0.01446	0.01552	0.02720
8.33%	0.08333	0.05749	0.04447	0.02584	0.03887
12.50%	0.08333	0.10913	0.08128	0.02580	0.00206
16.67%	0.08333	0.11965	0.12152	0.03632	0.03818
20.83%	0.16667	0.12932	0.16359	0.03734	0.00308
25.00%	0.16667	0.14136	0.20665	0.02531	0.03998
Total difference for this pair				0.16612	0.14937

Panel F: pair BD (Galambos)

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00783	0.00174	0.00783	0.00174
8.33%	0.04167	0.03707	0.00694	0.00460	0.03472
12.50%	0.04167	0.06548	0.01563	0.02382	0.02604
16.67%	0.04167	0.07776	0.02778	0.03609	0.01389
20.83%	0.04167	0.08322	0.04340	0.04155	0.00174
25.00%	0.04167	0.09841	0.06250	0.05674	0.02083
Total difference for this pair				0.17062	0.09896

Panel G: pair BE (Frank)

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00000	0.00015	0.00000	0.00015
8.33%	0.00000	0.00005	0.00073	0.00005	0.00073
12.50%	0.00000	0.00503	0.00197	0.00503	0.00197
16.67%	0.00000	0.00873	0.00418	0.00873	0.00418
20.83%	0.00000	0.01177	0.00781	0.01177	0.00781
25.00%	0.00000	0.02265	0.01347	0.02265	0.01347
Total difference for this pair				0.04822	0.02832

Panel H: pair CD (Clayton)

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.01400	0.00599	0.01400	0.00599
8.33%	0.04167	0.05045	0.01583	0.00878	0.02584
12.50%	0.04167	0.08756	0.02871	0.04590	0.01296
16.67%	0.04167	0.10206	0.04444	0.06039	0.00278
20.83%	0.04167	0.12320	0.06296	0.08153	0.02129
25.00%	0.08333	0.13932	0.08423	0.05598	0.00090
Total difference for this pair				0.26658	0.06975

Panel I: pair CE (Gaussian)

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00000	0.00006	0.00000	0.00006
8.33%	0.00000	0.00005	0.00060	0.00005	0.00060
12.50%	0.00000	0.00577	0.00231	0.00577	0.00231
16.67%	0.00000	0.01001	0.00590	0.01001	0.00590
20.83%	0.00000	0.01761	0.01209	0.01761	0.01209
25.00%	0.04167	0.03167	0.02155	0.00999	0.02011
Total difference for this pair				0.04342	0.04107

Panel J: pair DE (Joe)

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00002	0.00297	0.00002	0.00297
8.33%	0.00000	0.00923	0.01153	0.00923	0.01153
12.50%	0.00000	0.06704	0.02521	0.06704	0.02521
16.67%	0.00000	0.09455	0.04361	0.09455	0.04361
20.83%	0.04167	0.11596	0.06636	0.07429	0.02470
25.00%	0.12500	0.16891	0.09314	0.04391	0.03186
Total difference for this pair				0.28904	0.13988

Comparison between estimations of likelihood of joint extremely high default rates (normality vs. copulas). The survival copulas are informed in parenthesis after the names of the pairs and were estimated based on the right tails (above the 75th percentile) of the distributions of default rates.

3.7 CONCLUSIONS

Copula Theory has been employed in credit risk analyses but, to the best of our knowledge, this is the first investigation to present an empirical study of copulas for consumer loans. Moreover we test five copulas that are not typically considered in the literature and three of them (Galambos, Hüsler-Reiss, and Plackett)[23] were found to be representative of the dependence between some segments. Given that the main objective here is to find dependence structures (copulas) that yield more precise estimation of simultaneous high losses, another innovation was the inclusion of goodness-of-fit (GoF) tests based

[23] These three families were found when the GoF tests were based on the complete distributions of the default rates. Four “atypical” copulas resulted from estimations supported by the right tail of the default distributions: Galambos, Hüsler-Reiss, Joe, and Plackett.

exclusively on the right tails of the default distributions (along with the complete distributions, which is often done).

As for the usual strategy (considering the whole default distributions), among the ten segments investigated, eight present tail dependence[24] (i.e. higher association across extreme occurrences), from which five have upper-tail dependence, indicating that higher losses are more correlated. This suggests that, especially in downturns, the Financial Institution is subject to higher losses in the credit card portfolio than those assumed by traditional models. Only in one pair of segments the dependence is expressed by the Gaussian Copula (implicit in many models currently in use).

With regard to the alternative strategy (GoF based on the right tail of default distributions), seven pairs have tail dependence; six of them are right-tail dependent. This confirms the conclusion that most of the pairs tend to be more associated when default rates are higher (i.e. in unfavourable economic scenarios). None of the pairs is represented by the Gaussian Copula.

Although the Gaussian Copula (the basis of some traditional credit risk models) cannot be rejected for most of the pairs (since they cannot be statistically rejected due to the high p-values in the GoF tests), an important conclusion of this study is that the dependence across credit card loans can be better expressed by other copula families. This multiple representation for the dependence is likely due to the small amount of observations evaluated which is insufficient to generate a clear distinction among the copulas tested. However when the copulas are estimated considering only the right tail of the default distributions, the Gaussian Copula can be rejected in more pairs (particularly in analyses based on the Kendall's Transform GoF).

The limitations of traditional credit models in terms of estimation of joint extreme losses refer not only to the assumption of univariate losses' distributions but also to the treatment of the *joint* behaviour across defaults. In the credit card portfolio analysed here, we show some examples of segments (A, B, and C)

[24] Although the results for two of those pairs are not statistically significant.

with distributions statistically close to normality[25] whose dependence is far from the Gaussian Copula implicit in traditional credit risk models.

The comparison between joint extreme losses estimations derived from normality assumptions and copulas followed those two GoF strategies mentioned above. Evaluation using copulas yielded better results for 63% (70%) of the scenarios tested when the complete (right tails of) default distributions were used to estimate the copulas. When considering the average performance for each pair of segments, copula functions were more representative than bivariate normal distributions for nine pairs (ten pairs for the right-tail GoF approach). However, contrary to what is theoretically expected, the copula technique presented higher underestimation ratio of high losses (17% and 27% for the complete and right-tail distributions, respectively) than the Gaussian model (12%).

In sum, it seems that there is a trade-off between the two GoF approaches: the one based on the right-tail of default distributions selects copulas more representative of extreme defaults (which improves copula estimations of joint *high* defaults) at the expense of higher underestimation indices (which we want to avoid).

Nonetheless our conclusions are limited due to the short period covered by the dataset (24 months). Even though it includes some months with intense losses at the end of 2008 (the so-called “credit crunch”), which could result in a higher proportion of conjunct high default rates, it does not have enough observations to generate potential joint losses in the extremely upper tail of the distributions (98% or 99%, for example) where the biggest deficiency of traditional models seems to be[26]. Therefore a natural extension of this work is to apply the same procedure in a dataset covering a longer time horizon to verify the estimations

[25] Albeit the results of the test (Jarque-Bera) used were not significant (see Table 3.2).

[26] On the other hand, the number of observations used in this study is analogous to the length of datasets commonly available in many financial institutions. So, the limitations of this empirical analysis are likely the same to be found in practical applications of Copula Theory to banks' loan portfolios.

of joint events at extreme levels. In order to consolidate the use of copulas in consumer loans, the dependence structure and the probability of severe losses in other types of portfolios, e.g. mortgages and fixed term loans, should be assessed and compared to estimations from traditional models.

It is worth bearing in mind that the method used to estimate the copula parameters in this analysis assumes that the variables (default rates) do not present temporal dependence. In future studies, techniques that take serial correlation into account should be employed.

Also, due to the diversity of copulas found to represent the association between pairs of segments, it is interesting to search for a combination of copulas that represents the heterogeneous dependence across more than two segments together. This topic will be addressed in the next chapter.

CHAPTER 4
ESTIMATION OF JOINT DEFAULTS IN PORTFOLIOS WITH HETEROGENEOUS DEPENDENCE

"Most people are more comfortable with old problems than with new solutions."
(Anonymous)

4.1 INTRODUCTION

As shown in the previous chapter, pairs of credit segments present distinct dependence structures. This can be seen as evidence that the assumption of equal dependence for all pairs of loans or segments in a portfolio does not express the reality in financial institutions and therefore may lead to biased assessment of potential joint extreme losses[27].

To capture different relationships among loans and to still keep the advantage of Copula Theory concerning the identification of tail association, we use vine copulas, the basic idea of which is to decompose multivariate copulas into a cascade of bivariate copulas so that the estimation of higher-dimension dependence can be performed through the combination of relatively simple steps[28].

This approach was tested for the credit card portfolio data studied in Chapter 3. The credit segments were pooled into triplets and the dependence of each trio was estimated. The results showed a wide variety of structures that altered the conclusions from the pairwise analysis in some cases. The main drawback of bivariate examinations in the context of this thesis became evident in situations where none of the three bivariate copulas estimated separately for segments in a triplet pinpointed the right-tail dependence identified in the vine approach.

[27] Recall that the evaluation of joint losses in credit portfolios according to factor models presumes that the latent variables of all pairs of loans (and therefore their defaults) have the same correlation (represented by ρ in expression [2.4], Chapter 2).

[28] Puzanova et al. (2009) suggest three different techniques to estimate the risk in heterogeneous portfolios but the heterogeneity in that paper is related to the loans' size (i.e. the assumption of a portfolio composed of a large number of small loans is relaxed). There is no concern in terms of asymmetric dependence and the losses simulated to test the models are assumed to come from a multivariate normal distribution.

Once the best-fit copulas for each triplet were found, we estimated the likelihood of simultaneous high default rates following the vine construction and compared such estimates to calculations conditional on the assumption of normally-distributed losses. The evaluations based on vine copulas outperformed the normal method in 80% (65%) of the scenarios considered when the best-fit copulas were inferred from the right tail of (complete) default rates' distributions.

Unlike the use of bivariate copulas whose applications have spread into many financial fields, there are few applications of vine copulas in finance. Exceptions are Heinen and Valdesogo (2009), Aas et al. (2009), Aas and Berg (2009), and Maugis and Guegan (2010) who studied stocks' returns[29]. Therefore vine copulas have not been employed to model dependence across default rates in credit portfolios and this chapter aims to contribute to fill this gap in the literature.

Vine copulas are explained in the next section. Then the criteria to group the loans into triplets are described. In Section 4.4, the dependence for each triplet is estimated. Next, estimates of joint high losses derived from the best-fit vine copulas are compared to estimates based on the assumption of normally-distributed default rates. Section 4.6 concludes.

4.2 VINE COPULAS

4.2.1 Decomposition of conditional and multivariate distributions

Let $f(.)$ represent the density of a cumulative distribution function $F(.)$. From basic statistics (see, for instance, Casella and Berger, 2002), we know that

$$f(x_1 \mid x_2) = \frac{f(x_1, x_2)}{f(x_2)} \qquad [4.1]$$

[29] Aas et al. (2009) analysed bond indexes' returns as well. More general applications of vine copulas to dependence modelling are compiled in Kurowicka and Joe (2010).

for $f(x_2) > 0$ and where $f(x_1 \mid x_2)$ is the density of a variable X_1 evaluated at x_1 conditional on a variable X_2 evaluated at x_2. Therefore,

$$f(x_1, x_2) = f(x_2).f(x_1 \mid x_2)$$

For d random variables represented by a vector $\mathbf{X} = (X_1,..., X_d)$, the joint density function $f(x_1,..., x_d)$ can be decomposed into (see, e.g., Aas et al., 2009):

$$f(x_1,..., x_d) = f_d(x_d).f(x_{d-1} \mid x_d).f(x_{d-2} \mid x_{d-1}, x_d)...f(x_1 \mid x_2,..., x_d)$$

From the definition of copulas given in Chapter 2, we know that a joint distribution $F(.)$ may be expressed by means of a copula C:

$$F(x_1,..., x_d) = C(F_1(x_1),..., F_d(x_d))$$

Assuming that the joint distribution $F(.)$ is absolutely continuous and the d univariate distributions $F_1(x_1),..., F_d(x_d)$ are strictly increasing and continuous, we can apply the chain rule to derive $F(.)$ and get its density (see Aas et al., 2009):

$$f(x_1,..., x_d) = c_{1...d}(F_1(x_1),..., F_d(x_d)).f_1(x_1)...f_d(x_d) \qquad [4.2]$$

where $c_{1...d}$ is the d-variate copula density. Note that the copula captures the dependence amongst the variables and the marginals $f(.)$ express the shapes of the distributions. The bivariate case, for example, is:

$$f(x_1, x_2) = f_1(x_1).f_2(x_2).c_{12}(F_1(x_1), F_2(x_2)) \quad [4.3]$$

The combination of [4.1] and [4.3] yields the conditional density expressed in terms of the copula density:

$$f(x_1 \mid x_2) = f_1(x_1).c_{12}(F_1(x_1), F_2(x_2))$$

In general, each term involving conditional densities results in:

$$f(x \mid \mathbf{v}) = c_{xv|\mathbf{v}_{-j}}(F(x \mid \mathbf{v}_{-j}), F(v_j \mid \mathbf{v}_{-j})).f(x \mid \mathbf{v}_{-j})$$

where x is the conditioned variable, $\mathbf{v}$ is a vector of conditioning variables, $\mathbf{v}_{-j}$ is the vector $\mathbf{v}$ without the variable $\mathbf{v}_j$ and $c_{xv|\mathbf{v}_{-j}}$ is the density of the copula that links the conditional distribution functions $F(x|\mathbf{v}_{-j})$ and $F(v_j|\mathbf{v}_{-j})$. Such functions are calculated as in Joe (1996)[30]:

$$F(x \mid \mathbf{v}) = \frac{\partial C_{xv_j|\mathbf{v}_{-j}}(F(x \mid \mathbf{v}_{-j}), F(v_j \mid \mathbf{v}_{-j}))}{\partial F(v_j \mid \mathbf{v}_{-j})}$$

[30] A detailed proof of this formula is given in Czado (2010).

where $C_{xv_j|\mathbf{v}_{-j}}$ is a copula distribution function, $\mathbf{v}_j$ is a component of vector $\mathbf{v}$ and $\mathbf{v}_{-j}$ is the vector $\mathbf{v}$ excluding this component. When $\mathbf{v}$ is univariate, the conditional distribution becomes:

$$F(x \mid v) = C_{x|v}(F(x) \mid F(v)) = \frac{\partial C_{xv}(F(x), F(v))}{\partial F(v)} \qquad [4.4]$$

Since this chapter will associate default rates of three credit segments the trivariate version of [4.2] is of great interest here:

$$f_{x_1 x_2 | x_3}(x_1, x_2, x_3) = \qquad [4.5]$$

$$f_1(x_1).f_2(x_2).f_3(x_3).c_{13}(F_1(x_1), F_3(x_3)).c_{23}(F_2(x_2), F_3(x_3)).c_{12|3}(F(x_1 \mid x_3), F(x_2 \mid x_3))$$

However it is important to bear in mind that [4.5] is only one of the three possible representations of $f(x_1, x_2, x_3)$ given that permutations of x_1, x_2 and x_3 are likely to result in different copulas and conditional distributions $F(.|.)$. In this example, x_3 was randomly chosen as the conditioning variable (i.e. after the symbol "|" in the conditional distributions). The other two variables, x_1 and x_2, could also be selected and this would probably yield different values for the joint density $f(x_1, x_2, x_3)$. This essay does not have the objective of finding the arrangement of variables that leads to the best approximation to the dependence of an observed multivariate dataset. This topic is treated in Maugis and Guegan (2010) who suggested a method to identify the closest vine to the multivariate copula of a dataset.

Our derivation of [4.5] , following the concepts above, is shown in Apendix E. Decompositions of joint densities with up to five variables are presented in Bedford and Cooke (2001) and Aas et al. (2009).

4.2.2 Vines (a graphical introduction)

Formulas, such as [4.5], that combine several bivariate copulas to express the dependence among more than two variables may be defined with the help of acyclic graphs named *vines* that can be organised in the form of a nested set of trees (see Joe, 1996)[31]. Figure 4.1 displays one possible representation of the dependence among three variables X_1, X_2 and X_3. Each variable or combination of variables is named a *node*. Each *tree* is a diagram that expresses the links among nodes. The links between variables or pools of variables are called *edges* and each edge corresponds to a bivariate copula that will be a node in the subsequent tree.

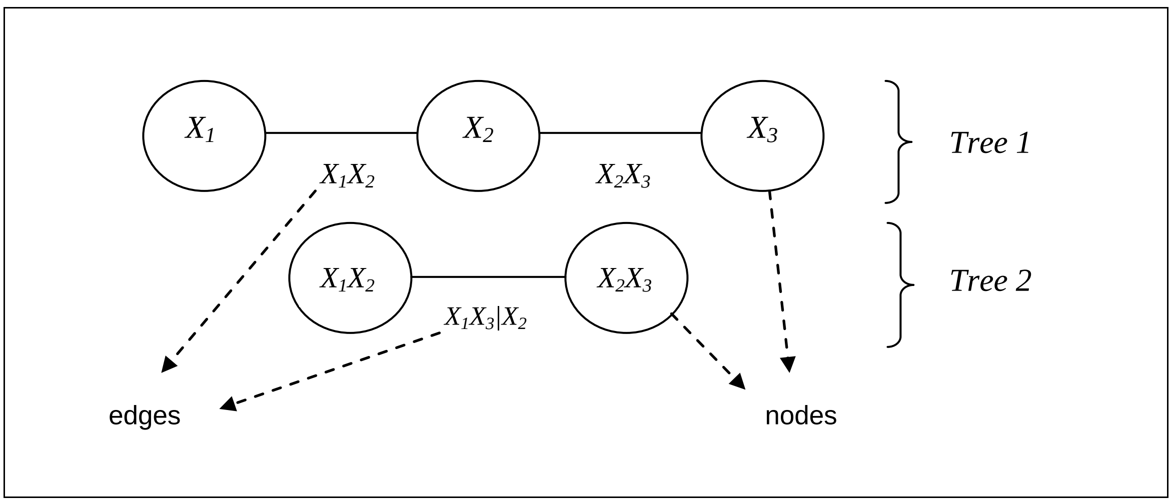

FIGURE 4.1: A dependence structure among three variables represented by two trees.

Tree 1 represents the connections across the variables (nodes) X_1, X_2 and X_3. The edges X_1X_2 and X_2X_3 in that tree indicate the dependence (copula) for

[31] We adopt here the term *vine* following Bedford and Cooke (2001, 2002) albeit it was not employed in the original study of Joe (1996). This method is also called Pair-Copula Construction (PCC) as in Aas et al. (2009) and Aas and Berg (2009).

each pair: $C(F(X_1), F(X_2))$ for the link between X_1 and X_2 and $C(F(X_2), F(X_3))$ for the link between X_2 and X_3.

Those two edges (X_1X_2 and X_2X_3) in *Tree 1* are the nodes of *Tree 2*. Given that X_2 is the common term in both nodes, it becomes the conditioning variable (placed after the symbol "|") in the edge $X_1X_3 | X_2$ which, in turn, symbolises the copula $C(F(X_1), F(X_3) | X_2) = C(F(X_1 | X_2), F(X_3 | X_2))$ referent to the dependence between X_1 and X_3 conditional on X_2. For the sake of simplicity, in the following figures, the edges will be written as combinations of variables (for instance, X_1X_2 or $X_1X_3 | X_2$) although, in fact, they represent their dependence (copulas).

If the composition of the trees is modified, the resultant nodes (and therefore the copulas) will change. In Figure 4.2, the positions of X_2 and X_3 are swapped when compared to Figure 4.1 and X_3 becomes the conditioning variable. Consequently, the final node is $X_1X_2 | X_3$ (instead of $X_1X_3 | X_2$ in Figure 4.1) and the related copula is $C(F(X_1 | X_3), F(X_2 | X_3))$.

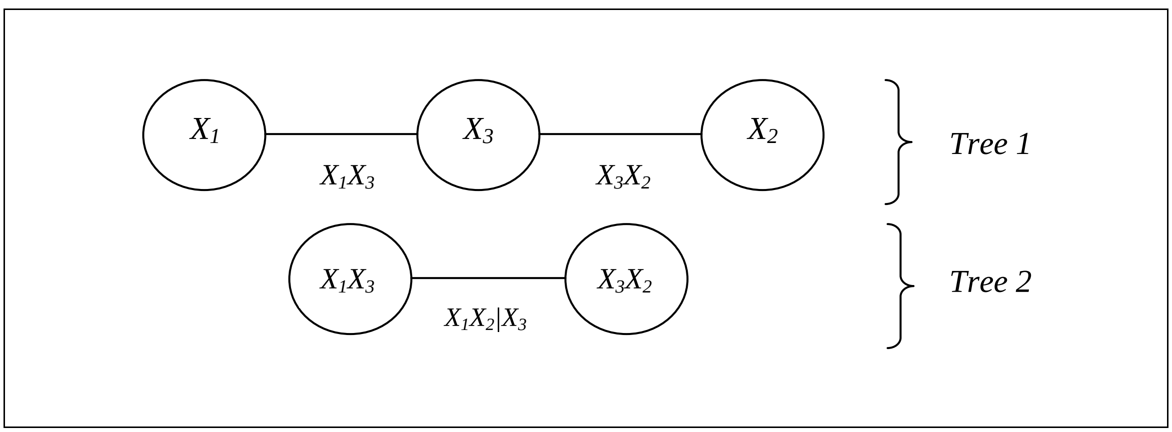

FIGURE 4.2: An alternative structure of dependence among three variables represented by two trees.

There are many ways to construct vines but we will focus only on *regular vines* suggested by Bedford and Cooke (2001, 2002). A vine is called regular if the number of its variables (d) is equal to the number of trees (T) plus one (i.e. $d = T + 1$).

Among the regular vines, D-vines and canonical vines are the most popular (see Kurowicka and Cooke, 2002 and Aas et al., 2009). They differ from each other with respect to the way of decomposing the density. Whilst in D-vines no node in any tree is connected to more than two edges (see Figure 4.3), in canonical vines, each tree has a unique node that is connected to all other nodes (see Figure 4.4 where, in tree T_1, for example, X_1 is linked to all other four nodes, X_2, X_3, X_4 and X_5). The trivariate analysis is a special case where the D-vine and the canonical vine result in the same expression for the density (Aas et al., 2009).

The D-vine depicted in Figure 4.3 consists of four trees T_q $(q = 1,...,4)$ where tree T_q has $6-q$ nodes and $5-q$ edges. Each edge corresponds to a bivariate copula that will be part of the decomposition of the multivariate density. In general, a whole decomposition is defined by $d(d-1)/2$ edges and the marginal density of each variable. In the example above, there are 10 edges, so the density expression will have 10 bivariate copulas (four unconditional and six conditional):

$$f(x_1, x_2, x_3, x_4, x_5) = f_1 . f_2 . f_3 . f_4 . f_5 . c_{12} . c_{23} . c_{34} . c_{45} . c_{12|3} . c_{24|3} . c_{35|4} . c_{14|23} . c_{25|34} . c_{15|234}$$

[4.6]

where $f(x_1, x_2, x_3, x_4, x_5)$ is the joint density of five variables evaluated at those respective values and the remaining notation is simplified for convenience: f_i is the marginal density of X_i at x_i, c_{ij} is the density copula of $F(x_i)$ and $F(x_j)$, and $c_{ij|k}$ is the conditional density copula of $(F(x_i \mid x_k), F(x_j \mid x_k))$.

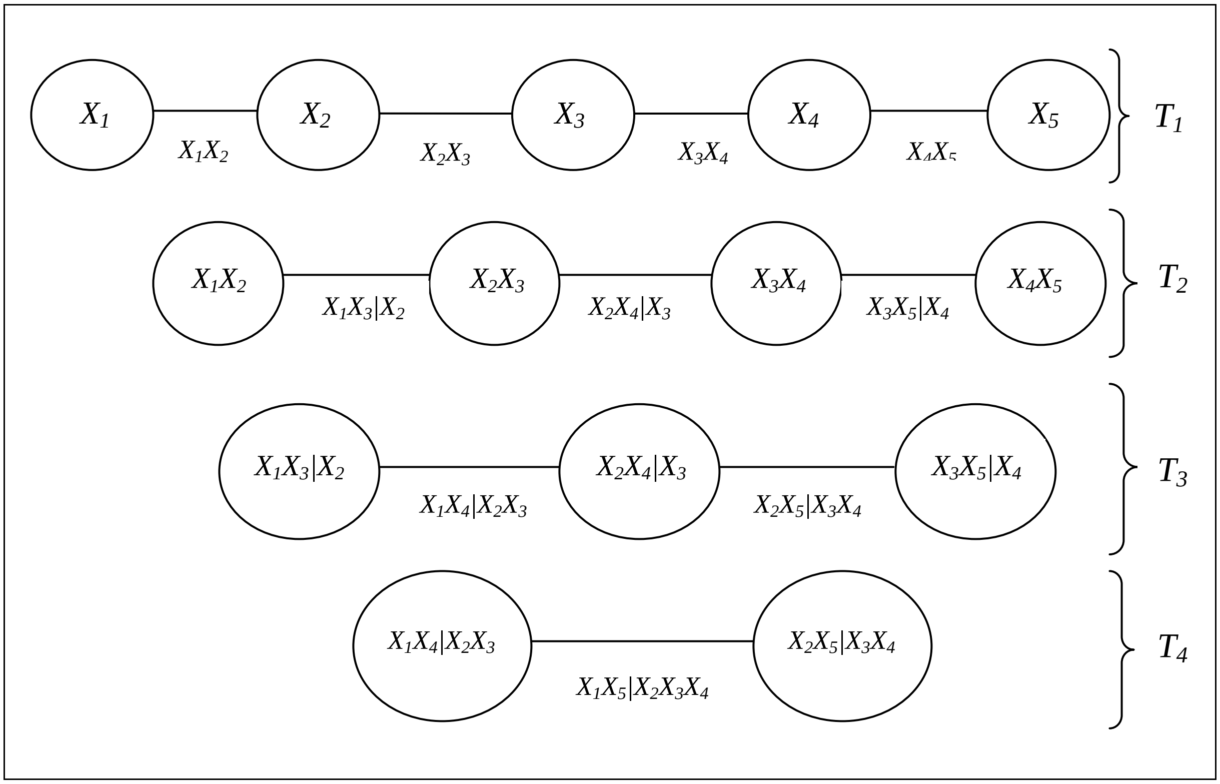

FIGURE 4.3: D-vine depicting the dependence among five variables.

Figure 4.4 shows a canonical vine with five variables. The multivariate density depicted in that diagram will be:

$$f(x_1, x_2, x_3, x_4, x_5) = f_1 . f_2 . f_3 . f_4 . f_5 . c_{12} . c_{13} . c_{14} . c_{15} . c_{23|1} . c_{24|1} . c_{25|1} . c_{34|12} . c_{25|34} . c_{45|123}$$

where the same notation of [4.6] applies. Notice that most of the (bivariate) copulas are not the same for the D-vine and the canonical vine in [4.6].

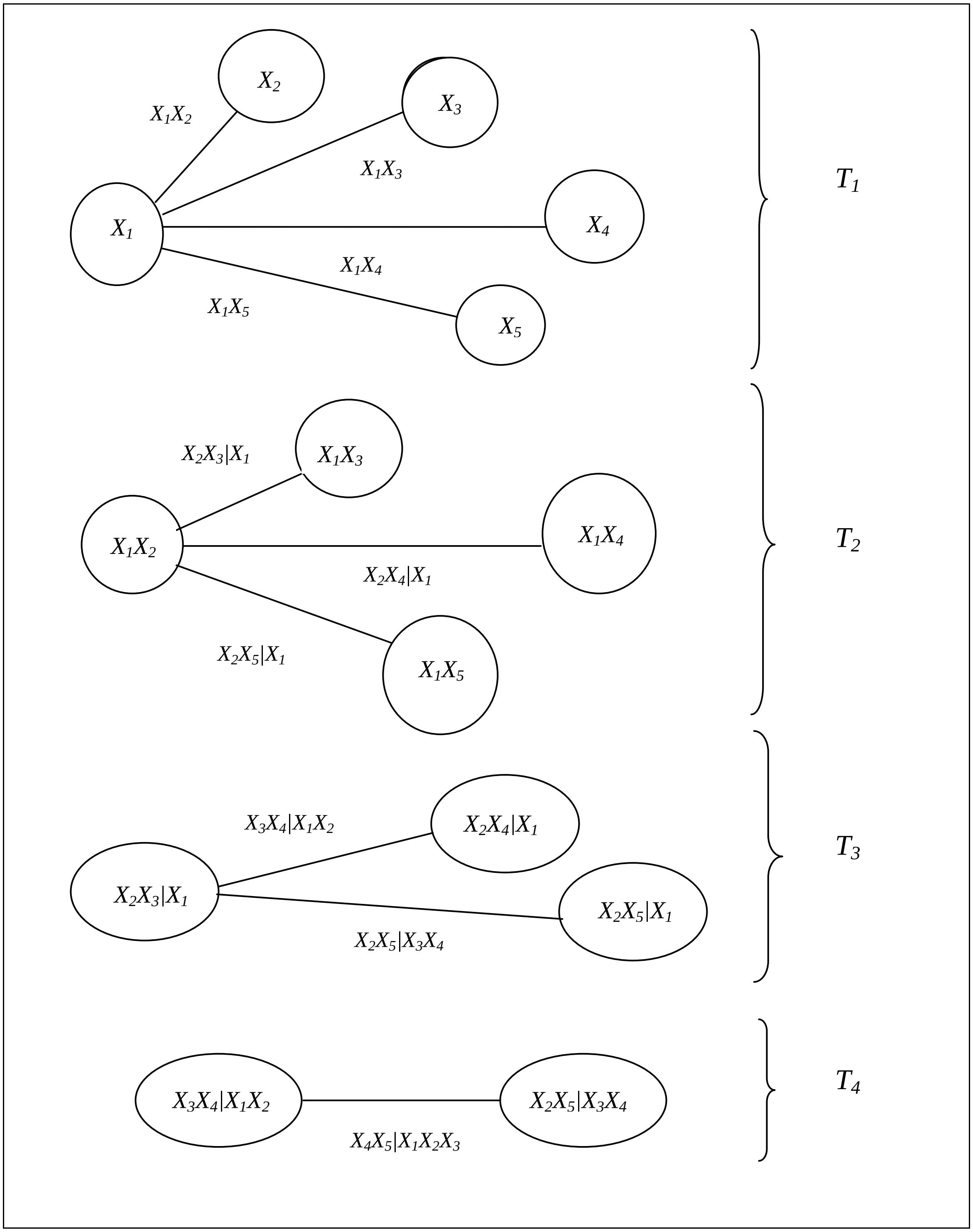

FIGURE 4.4: Canonical vine depicting the dependence among five variables.

An example of a regular vine that is neither a D-vine nor a canonical vine could start with the tree illustrated in Figure 4.5. It is not a D-vine because the node X_3 is connected to more than one node and it is not a canonical vine because none of the nodes is connected to all other nodes.

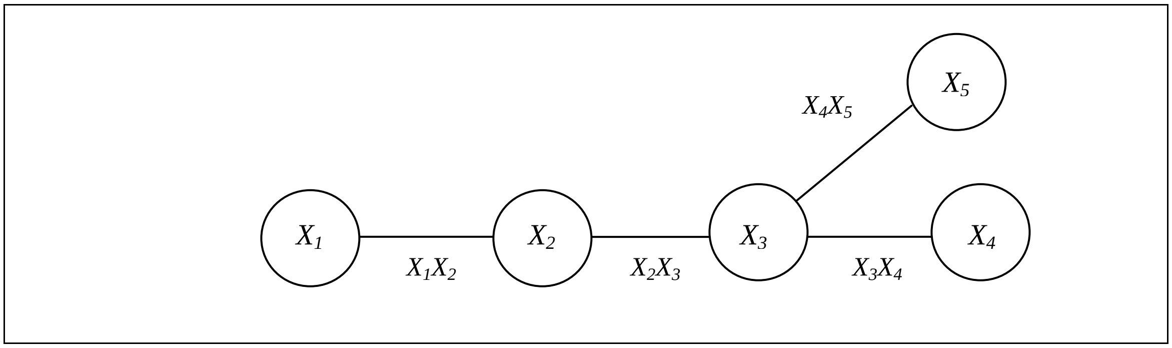

FIGURE 4.5: A regular vine that is neither a D-vine nor a canonical vine.

4.2.3 Nested (Archimedean) Copulas

High-dimension dependence can also be expressed by means of nested copulas (copulas of the form $C(u,v)$ in which at least one of the terms u and v is another copula). This technique is restricted to Archimedean copulas[32] since this class is the only one to present the associative property that allows such a combination.

Figures 4.6 and 4.7 exhibit two possible nested copulas linking three variables (each). To ease the comparison between vine and nested copulas, the notation used in Section 4.2.2 is kept such that each combination of variables represents the related copula (for instance, in Figure 4.6, X_1X_2 and $(X_1X_2)X_3$ stand for the copulas $C_{12}(F(X_1),F(X_2))$ and $C_{12,3}[C_{12}(F(X_1),F(X_2)),F(X_3)]$, respectively, where the subscripts were added to the copula notations in order to distinguish the bivariate copula from the nested copula which may belong to the different families of Archimedean copulas).

[32] The general form of Archimedean copulas was shown in Chapter 2, Section 2.3.2.2.

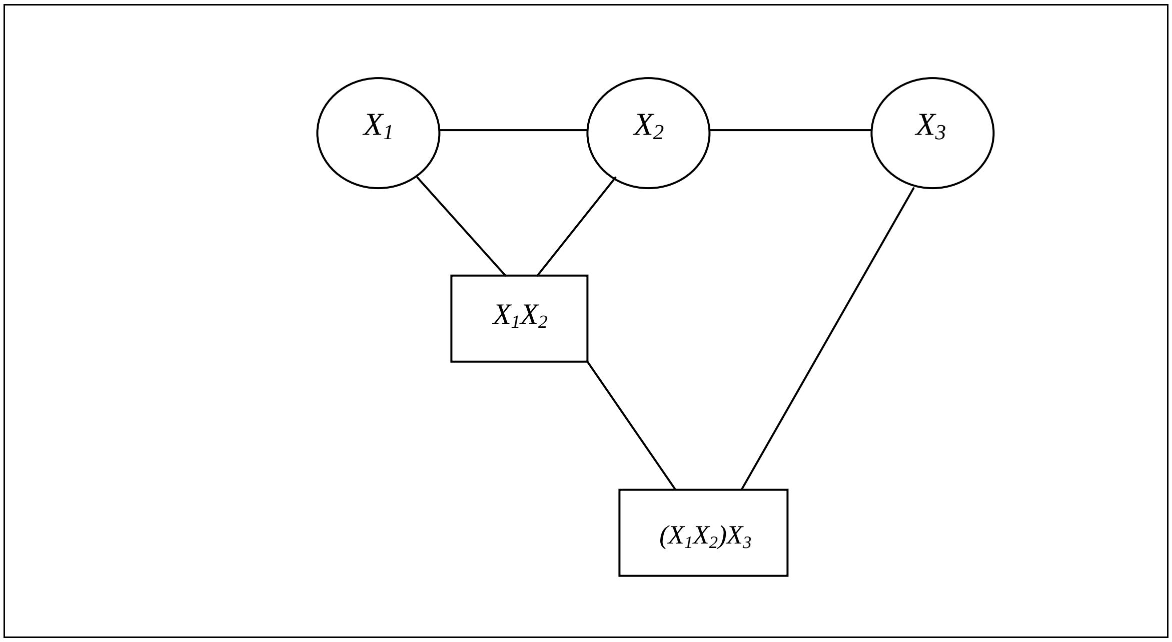

FIGURE 4.6: A nested copula representing the dependence across three variables.

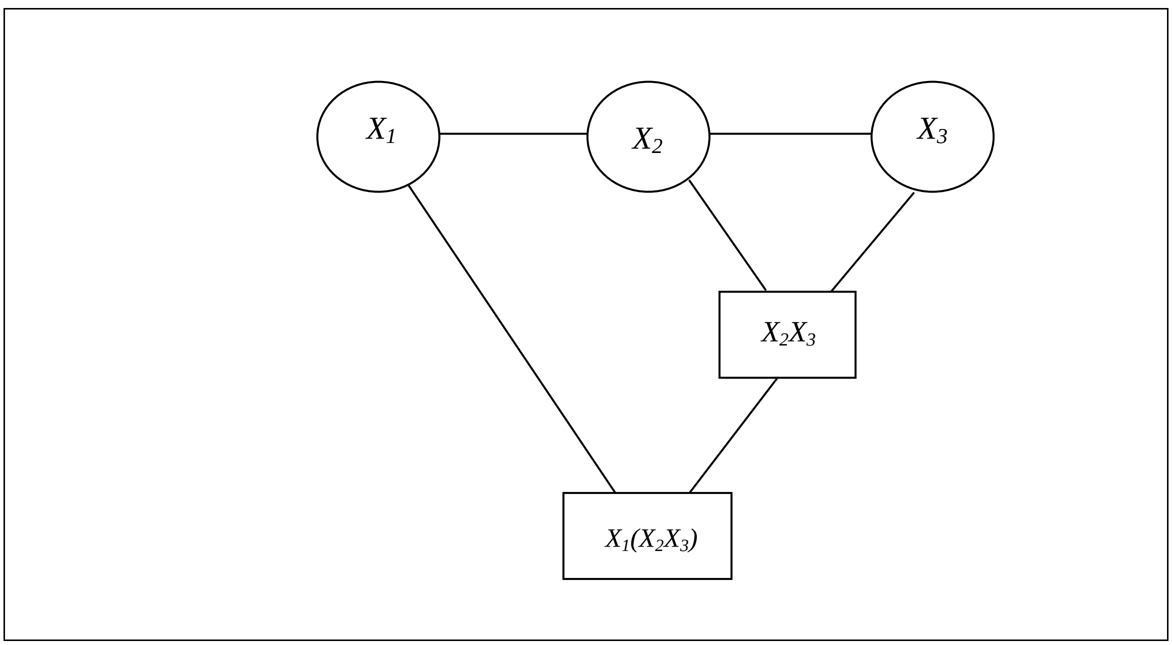

FIGURE 4.7: An alternative nested copula to represent the dependence across three variables.

Bear in mind that the aforementioned nested construction is distinct from a multivariate Archimedean copula of the form $C(F(X_1), F(X_2), F(X_3))$ in which one Archimedean copula family C connects all variables (three, in this case) through the *same* structure (see, e.g., Embrechts et al., 2003 and Nelsen, 2006, Chapter 5). Consequently, nested representations are more flexible than the multivariate versions of copula families given that the former can combine diverse relationships in different pairs of variables and therefore capture more complex joint dependence. The copula $C(F(X_1), F(X_2), F(X_3))$, for example, indicates that all the variables studied (X_1, X_2 and X_3) follow the same dependence structure while the nested copula $C_{12,3}[C_{12}(F(X_1), F(X_2)), F(X_3)]$ may express a more complex and realistic structure if $C_{12,3}$ and C_{12} represent different Archimedean families (C_{12} = Clayton and $C_{12,3}$ = Frank, for instance, which means that X_1 and X_2 are left-tail dependent and that the relationship between this pair and the other variable, X_3, is symmetric without tail dependence).

Even though the initial structure of the three variables in Figures 4.1, 4.6 and 4.7 is the same, neither the nested copula in Figure 4.6, $C_{12,3}[C_{12}(F(X_1), F(X_2)), F(X_3)]$, nor the nested copula in Figure 4.7, $C_{1,23}[F(X_1), C_{23}(F(X_2), F(X_3))]$, is equivalent to the vine copula of Figure 4.1, $C(F(X_1 \mid X_2), F(X_3 \mid X_2))$.

Aas and Berg (2009) found out that vine copulas (pair-copula constructions) outperformed nested copulas in relation to the best fit to two multivariate datasets checked (one of which was pertaining to equity returns). Furthermore, the authors highlighted that, although both methods rely on the same principle (decomposition of multivariate dependencies into cascades of bivariate copulas), vine copulas are not restricted to the Archimedean class and can generate more potential structures; for d variables, $d(d-1)/2$ vines can be formed whilst nested Archimedean copulas originate only $d - 1$ combinations. This latter aspect is important given that the greater number of constructions from the vine approach gives more options to find the best-fit dependence for high-dimension data. In the trivariate case ($d = 3$), vines and Archimedeans have three and two potential combinations respectively and as the number of variables goes up, the advantage of vines becomes evident: for $d = 4$, vines result in six combinations and nested Archimedean copulas in three; for $d = 5$, vines give 10 different structures against four from nested associations.

Due to the advantages of vine constructions over nested Archimedean copulas, our analysis will use the former method. Readers interested in more details about nested copulas should refer to, e.g., Whelan (2004), Morillas (2005), and Savu and Trede (2006).

4.2.4 Estimation of joint cumulative distribution functions

The literature on vine copulas usually analyses the decomposition of joint dependence in terms of densities. In this study, we are particularly interested in joint *cumulative* distributions which can be found through the integration of the respective densities. Cumulative versions of joint densities are given in Joe (1996, 1997). For three variables X_1, X_2 and X_3 evaluated at x_1, x_2 and x_3 respectively, the joint cumulative distribution is:

$$F(x_1, x_2, x_3) = \int_{-\infty}^{x_3} C_{12|3}(F(x_1 \mid X_3), F(x_2 \mid X_3)).F_3(dX_3) \qquad [4.7]$$

where $F(.)$ is a distribution function and $C_{12|3}$ is the copula that links the two distributions $F(.|.)$ conditional on X_3. An alternative way to estimate [4.7] is presented in Joe et al. (2010):

$$C(F(x_1), F(x_2), F(x_3)) = \int_0^{F(x_3)} C_{12|3}[C_{1|3}(F(x_1) \mid F(X_3)), C_{2|3}(F(x_2) \mid F(X_3))]dF(X_3) =$$

$$= C_{12|3}[C_{13}(F(x_1), F(x_3)), C_{23}(F(x_2), F(x_3))] \qquad [4.8]$$

where the trivariate copula $C(F(x_1), F(x_2), F(x_3))$ that returns the joint probability $\Pr[X_1 \le x_1, X_2 \le x_2, X_3 \le x_3] = F(x_1, x_2, x_3)$ is factorised into a vine structure composed of three copulas: $C_{1|3}$ and $C_{2|3}$ (copulas of $F(x_1)$ and $F(x_2)$ respectively, conditional on $F(X_3)$ such that $0 \le X_3 \le x_3$) and $C_{12|3}$ (a copula that links those conditional associations and indicates the dependence between X_1 and X_2 given X_3). Note that, when integrated with respect to the conditioning variable X_3, $C_{1|3}$ and $C_{2|3}$ become the unconditional copulas C_{13} and C_{23} that associate the pairs $(F(x_1), F(x_3))$ and $(F(x_2), F(x_3))$, respectively. So, the trivariate *cumulative* distribution function represented in [4.8] is given by one copula ($C_{12|3}$) that connects two other copulas (C_{13} and C_{23}) that have a common term ($F(x_3)$) while the trivariate *densities* shown in Figures 4.1 and 4.2 are related to copulas that link conditional distributions.

The likelihood of simultaneous occurrences above specific values $\bar{F}(x_1, x_2, x_3) = \Pr[X_1 > x_1, X_2 > x_2, X_3 > x_3]$ may be found through the trivariate survival copula inferred from [4.8]:

$$\hat{C}(\bar{F}(x_1), \bar{F}(x_2), \bar{F}(x_3)) = \int_{F(x_3)}^{1} \hat{C}_{12|3}[\hat{C}_{1|3}(\bar{F}(x_1) \mid \bar{F}(X_3)), \hat{C}_{2|3}(\bar{F}(x_2) \mid \bar{F}(X_3))]d\bar{F}(X_3) =$$

$$= \hat{C}_{12|3}[\hat{C}_{13}(\bar{F}(x_1), \bar{F}(x_3)), \hat{C}_{23}(\bar{F}(x_2), \bar{F}(x_3))] \qquad [4.9]$$

where $\bar{F}(.) = 1 - F(.)$ and $\hat{C}$ stands for a survival copula (explained in Chapter 2). The remaining notation follows [4.8].

4.3 SEGMENTATION CRITERIA

This trivariate analysis uses the same dataset employed in the bivariate analysis (Chapter 3). The credit segments are as defined in the previous chapter such that "A" is the least risky and "E" is the riskiest. In this chapter, they were grouped into triplets and, since there are five segments, ten trios were formed.

As said before, a joint density with three variables (like [4.5]) can be factorised in three different ways and each of them has a different conditioning variable. Consequently, the same applies to [4.7], [4.8] and [4.9]. In this empirical analysis, the conditioning variable was always represented by the default rates of the riskiest segment in the triplet. For example, the estimation of potential joint extreme default rates in trio BCD, x_B, x_C and x_D, based on [4.9], used survival copulas of survival distributions of default rates in segments B and C conditional on default rates of the riskiest segment D, $\hat{C}_{B|D}(\overline{F}(x_B) \mid \overline{F}(X_D))$ and $\hat{C}_{C|D}(\overline{F}(x_C) \mid \overline{F}(X_D))$, respectively, where $\hat{C}$ indicates a survival copula, $\overline{F}(.)$ is a survival distribution and X_D is the default rate of segment D to be evaluated from x_D to 1.

Obviously, the triplets could be arranged in different ways: either conditional on the least risky segment or on the mid-risky one. We opted for setting up the dependence conditional on the riskiest segment because this corresponds to the most conservative scenario[33].

[33] This is due to the fact that, compared to other segments, the riskiest segment has higher idiosyncratic risk (see Das and Geng, 2006) and is more prone to reach its highest levels of default even when the economy is not in the worst state. Thus, studying the trivariate dependence when the riskiest segment presents its highest losses increases the chance of identifying possible conjunct extreme losses in economic situations that do not characterise downturns.

4.4 ESTIMATION OF DEPENDENCE STRUCTURE AMONG THREE SEGMENTS IN A CREDIT CARD PORTFOLIO

4.4.1 Approach to estimate best-fit copulas and their respective parameters

The type of dependence structure investigated here is compatible with D-vines structures (defined in Section 4.2). However since we associate three credit segments, and a D-vine and a canonical vine for the same three variables are identical to each other, the resultant vine copulas in this empirical research are equivalent to both types of vines.

Canonical vines for dimensions equal to or greater than four are interesting, for example, to study dependence among default rates in several credit segments conditional on a (macroeconomic) factor. This problem follows the structure presented in Figure 4.4 where the node labelled as X_1 would stand for the factor that impacts all credit segments (variables X_2, X_3, X_4 and X_5 in that case).

The estimation of the dependence structure within each triplet defined in Section 4.3 was based on [4.8]. In order to find the values necessary to calculate that expression, the following steps were adopted:

(i) Denote the default rate of a segment i as x_i. The conditional distributions $F(x_i \mid x_j) = C_{i|j}(F(x_i) \mid F(x_j))$ were calculated according to [4.4] and the bivariate copulas estimated in Chapter 3 were used to express the relationship between the credit segments. For example, the level of default rates of segment A conditional on default rates of segment E, $F(x_A \mid x_E) = C_{A|E}(F(x_A) \mid F(x_E))$, was estimated as the first derivative of the copula that represents their dependence (Student t, found in Chapter 3 based on the Empirical Copula goodness-of-fit test, taking into account the whole default rate distributions) with respect to $F(x_E)$, where $F(.)$ is the distribution of default rates;

(ii) The copula $C_{ij|k}$ that links the dependence between default rates of segments i and j given the default rates of segment k was estimated according to Empirical Copula goodness-of-fit test which was found in the literature to be the most robust (see Chapter 3). That is, the copulas $C_{i|k}$ and $C_{j|k}$ calculated in step (i) were used as margins to estimate the copula $C_{ij|k}$. The candidate copulas were the same ten families considered in Chapter 3 (see Table 3.3). The joint dependence within trio ADE, for instance, is represented by a vine structure in which the Clayton Copula $(C_{AD|E})$ connects the dependence of A conditional on E ($C_{A|E}$, a Student t Copula found in Chapter 3), to the dependence of D conditional on E ($C_{D|E}$, Plackett Copula estimated in Chapter 3). The results for all triplets will be presented ahead;

(iii) After finding the best-fit copula $C_{ij|k}$ that links the conditional distributions, the parameters of the three copulas involved ($C_{i|k}$, $C_{j|k}$ and $C_{ij|k}$) were simultaneously calculated by means of the maximum pseudo-likelihood method proposed by Aas et al. (2009). For D-vines in general, with d variables, N observations, density copulas c and distribution functions $F(.)$, the log-likelihood function is:

$$\sum_{r=1}^{d-1}\sum_{s=1}^{d-r}\sum_{n=1}^{N}\log[c_{s,s+r|s+1,\ldots,s+r-1}\{F(x_{s,n} \mid x_{s+1,n},\ldots,x_{s+r-1,n}),\, F(x_{s+r,n} \mid x_{s+r,n},\ldots,x_{s+r-1,n})\}]$$

for $1 \le r \le d-1$, $1 \le s \le d-r$ (where r and s are subscripts used to indicate the variables), and observations $1 \le n \le N$, meaning that the conditioning terms of each density copula $c_{s,s+r|.}(F(x_{s,n} \mid .), F(x_{s+r,n} \mid .))$ are all $x_{q,n}$ such that $s \le q \le s+r$. When $d = 3$, the joint log-likelihood function of default rates x_i, x_j and x_k, maximised with respect to the three parameters $\boldsymbol{\theta} = (\theta_{ik}, \theta_{jk}, \theta_{ij|k})$, is:

$$\sum_{n=1}^{N} \{\log[c_{ik}(F(x_{i,n}), F(x_{k,n}); \theta_{ik})] + \log[c_{jk}(F(x_{j,n}), F(x_{k,n}); \theta_{jk})] + \log[c_{ij|k}(F(x_{i,n}) \mid F(x_{k,n}), F(x_{j,n}) \mid F(x_{k,n}); \theta_{ij|k})]\} \quad [4.10]$$

See Aas et al. (2009, Appendix B) for a didactic example.

The procedures from (i) to (iii) were conducted twice. In the first case, the copulas were estimated based on the complete distributions of default rates and then the right tails of those distributions (above their 75th percentile) were used.

4.4.2 Estimation based on the complete default rate distributions

Table 4.1 shows the best-fit copulas for the ten triplets and their related parameters estimated in accordance with [4.10]. Table 4.2 reports the shape of the dependence structure implied by those copulas.

Table 4.1: Best-fit copulas* for trios of credit segments according to the complete distributions of default rates

Trio	Dependence ($i \mid k, j \mid k$)		Dependence ($i \mid k$)		Dependence ($j \mid k$)	
ijk	**Copula**	**Parameter**	**Copula**	**Parameter**	**Copula**	**Parameter**
ABC	Plackett	0.8874	Clayton	3.0355	Clayton	11.3341
ABD	Frank	-0.3062	Galambos	0.0100	Hüsler	0.6412
ABE	Student t	0.7667	Student t	-0.4430	Student t	-0.5059
ACD	Frank	-0.4420	Galambos	0.0113	Hüsler	0.6700
ACE	Frank	0.1696	Student t	-0.4430	Gaussian	-0.4323
ADE	Clayton	0.0551	Student t	-0.4430	Plackett	3.5806
BCD	Gaussian	1.0000	Hüsler	0.6736	Hüsler	0.6598
BCE	Gaussian	0.4012	Student t	-0.5059	Gaussian	-0.4323
BDE	Student t	0.0515	Student t	-0.5059	Plackett	3.5806
CDE	Frank	4.0991	Gaussian	-0.4323	Plackett	3.5806

* Estimation based on the Empirical Copula goodness-of-fit test and whole distributions of default rates.

Table 4.2: Dependence characteristics of the best-fit copulas estimated from the complete distributions of default rates

Trio	Dependence ($i \mid k, j \mid k$)		Dependence ($i \mid k$)		Dependence ($j \mid k$)	
ijk	Symmetry	Tail dependence	Symmetry	Tail dependence	Symmetry	Tail dependence
ABC	Yes	No	No	Yes, left	No	Yes, left
ABD	Yes	No	No	Yes, right	No	Yes, right
ABE	Yes	Yes	Yes	Yes	Yes	Yes
ACD	Yes	No	No	Yes, right	No	Yes, right
ACE	Yes	No	Yes	Yes	Yes	No
ADE	No	Yes, left	Yes	Yes	Yes	No
BCD	Yes	No	No	Yes, right	No	Yes, right
BCE	Yes	No	Yes	Yes	Yes	No
BDE	Yes	Yes	Yes	yes	Yes	No
CDE	Yes	No	Yes	no	Yes	No

Both tables corroborate the conclusion in Chapter 3 pertaining to the portfolio's heterogeneity vis-à-vis the dependence structure. It becomes clear that the vine approach is flexible and able to capture many potential combinations of dependence between pairs of segments when the analysis focuses on larger groups (trios in this study).

There is no general pattern of association between the conditional distributions – represented by ($i \mid k$) and ($j \mid k$) in Tables 4.1 and 4.2 – and the copula that links them – expressed as ($i \mid k, j \mid k$) in those tables. For example, the joint copulas $C_{ij|k}$ in triplets ABC, BCD, BCE and CDE have the same feature (symmetric dependence without strong tail relationship) but their conditional distributions present four distinct combinations. Also, the existence of two symmetric conditional distributions (see triplets ABE, ACE, ADE, BCE, BDE and CDE) does not guarantee the symmetry of the joint copula.

Most (nine out of ten) of the triplets present symmetric joint copulas (Plackett, Frank, Student t or Gaussian) and seven of them do not have tail dependence. This suggests that the borrower's risk will not be excessively high in downturns. Two cases (ABD and ACD) confirm this: they have two right-tail dependent conditional distributions (meaning that higher default rates are more associated) linked by a symmetric copula without tail dependence and negative parameter which offsets the potential increase of losses in scenarios of high default levels[34].

Another favourable characteristic of this credit card portfolio is the fact that the riskiest triplet (CDE) is represented by three symmetric copulas without tail dependence (Frank, Gaussian and Plackett). This avoids disproportionate increments of losses in downturns since the loans with the highest propensity to default do not have stronger association in those adverse circumstances.

4.4.3 Estimation based on the upper tails of default rate distributions

The best-fit copulas were also estimated by restricting the goodness-of-fit test to the right tail (above the 75th percentile) of the default's distributions (as explained in Chapter 3, Section 3.5.2). The copulas and their parameters calculated according to [4.10] are reported in Table 4.3. To help with the interpretation of the dependence structures, Table 4.4 details the features of the copulas found.

Most of the cases (six triplets) have a joint copula $C_{ij|k}$ with right-tail association although four of them (ABD, ABE, ACD and ACE) have parameters very close to zero, which means that such conditional distributions are virtually independent. This is beneficial to the Bank given that it reduces the possibility of simultaneous high losses. Another four triplets (ABC, ADE, BDE and CDE) are represented by symmetric copulas without tail dependence. Thus, this evaluation based on vine copulas estimated according to the right tail of the distributions reveals the benefits from the heterogeneous dependence in the portfolio.

[34] BCD also has this property but its joint copula (Gaussian) has parameter equal to 1 which indicates complete dependence and therefore does not contribute to diversify the risk of the two default's distributions associated.

Table 4.3: Best-fit copulas* for trios of credit segments according to right-hand tails of default rates' distributions

Trio	Dependence ($i \mid k, j \mid k$)		Dependence ($i \mid k$)		Dependence ($j \mid k$)	
ijk	**Copula**	**Parameter**	**Copula**	**Parameter**	**Copula**	**Parameter**
ABC	FGM**	-0.3349	Joe	1.2685	Clayton	11.3341
ABD	Galambos	0.0000	Galambos	0.0726	Gumbel	171.5563
ABE	Galambos	0.0128	Frank	-3.3525	Frank	-3.9902
ACD	Galambos	0.0120	Galambos	0.0131	Joe	1.1269
ACE	Galambos	0.0127	Frank	-3.3525	Plackett	0.0832
ADE	Gaussian	0.0494	Frank	-3.3525	Husler	0.1259
BCD	Joe	3.1704	Gumbel	1.1175	Joe	1.1269
BCE	Joe	42.8336	Frank	-3.9902	Plackett	0.0832
BDE	Gaussian	0.0884	Frank	-3.9902	Husler	0.1261
CDE	Gaussian	0.2074	Plackett	0.0832	Husler	0.0058

* Estimation based on the Empirical Copula goodness-of-fit test and right-hand tails of default rates' distributions.
** FGM stands for Farlie-Gumbel-Morgenstern.

Table 4.4: Dependence characteristics of best-fit copulas estimated from the right-hand tails of default rates' distributions

Trio	Dependence ($i \mid k, j \mid k$)		Dependence ($i \mid k$)		Dependence ($j \mid k$)	
ijk	**Symmetry**	**Tail dependence**	**Symmetry**	**Tail dependence**	**Symmetry**	**Tail dependence**
ABC	Yes	No	No	Yes, right	No	Yes, left
ABD	No	Yes, right	No	Yes, right	No	Yes, right
ABE	No	Yes, right	Yes	No	Yes	No
ACD	No	Yes, right	No	Yes, right	No	Yes, right
ACE	No	Yes, right	Yes	No	Yes	No
ADE	Yes	No	Yes	No	No	Yes, right
BCD	No	Yes, right	No	Yes, right	No	Yes, right
BCE	No	Yes, right	Yes	No	Yes	No
BDE	Yes	No	Yes	No	No	Yes, right
CDE	Yes	No	Yes	No	No	Yes, right

On the other hand, this study also shows three instances (ABE, ACE and BCE) in which the pairs of conditional distributions have weak dependence at extreme points but the copulas that connect those distributions have upper-tail association. Thus the broader view of the portfolio indicates that it is riskier than supposed when we are restricted to the pairwise analysis. BCE illustrates this difference. According to the isolated use of bivariate copulas (see Table 3.5 in Chapter 3), pairs BE and CE are negatively associated (with no tail dependence) and the pair BC has left tail dependence (i.e. lower default rates more correlated). So there is not any evidence of a stronger connection among high default rates of B, C and E. However when those three segments are evaluated together, it is found that the default rates of B and C tend to be more associated once the defaults of the riskiest segment E reaches their highest levels. That is, the relationship between the distributions of B and C given E has upper-tail dependence whilst the unconditional distributions of B and C are lower-tail dependent[35].

Three groups (ABD, ACD and BCD) present the most alarming combination for the Bank: three right-tail dependent copulas indicating that defaults have tendency to become more associated in downturns. Nonetheless only the parameter for the joint copula of the last trio (BCD) is considerably above the (minimum) value that indicates independence.

ADE, BDE and CDE are intermediary combinations from the risk standpoint inasmuch as they present mixtures of right-tail dependent copulas and copulas without tail dependence.

4.5 ESTIMATION OF JOINT HIGH DEFAULTS IN THREE SEGMENTS: COMPARISON BETWEEN TRADITIONAL METHODS AND VINE COPULAS

A practical effect of the diversified dependence structure found in the previous section is the possible misevaluation of joint defaults in three segments when normal distributions (univariate and trivariate) are assumed. The use of vine copulas is expected to yield better approximations to estimations of simultaneous credit losses since such a technique is more flexible and can capture different combinations of dependence between pairs of segments that form loan portfolios.

[35] Trios ABE and ACE also have this property but, as said before, the parameters of the copulas that link the conditional distributions are very close to the respective values that indicate independence.

For a triplet involving segments I, J and K, the estimation of joint default rates, X_I, X_J and X_K, above particular values, x_I, x_J and x_K respectively, based on the assumption of normality is given by:

$$\Pr(X_I > x_I, X_J > x_J, X_K > x_K) =$$
$$1 - \Phi(x_I) - \Phi(x_J) - \Phi(x_K) + \Phi(x_I, x_J) + \Phi(x_I, x_K) + \Phi(x_J, x_K) - \Phi(x_I, x_J, x_K)$$

where Φ denotes the *cdf* of a normal distribution.

As in Chapter 3, estimations following Copula Theory will use survival copulas of the default rates, such that:

$$\Pr(X_I > x_I, X_J > x_J, X_K > x_K) = \hat{C}[1 - F_I(x_I), 1 - F_J(x_J), 1 - F_K(x_K)]$$

where $\hat{C}$ is a survival copula, i.e. links “survival ranks”: $1 - F(.)$; F_I, F_J and F_K are the *cdfs* of the (unknown) distributions of default rates X_I, X_J and X_K in that order.

The survival copulas applied in this calculation were estimated according to the procedures described in Section 4.4.1, the only difference being the replacement of $F(.)$ with $1 - F(.)$.

The proportions of default levels are the same defined in the pairwise analysis (Chapter 3): 4.17%, 8.33%, 12.50%, 16.67%, 20.83% and 25% (which we define as *high* losses). Given that the survival copulas were estimated with respect to whole default’s distributions and their right tails, results based on both cases are presented below.

4.5.1 Survival copulas estimated considering the complete distributions of default rates

Table 4.5 displays the comparison between estimates derived from normality assumptions and from vine copulas when the survival copulas were found as the best fit to the complete default's distributions. The first column refers to the *proportion* of the extreme defaults (not the default rates). The second column is related to the credit card portfolio analysed and displays the ratio of concurrent default rates above the mentioned levels. Columns "Normal" and "Copula" show the estimations from normality and vine copula approaches. The two last columns give the absolute difference between both estimations and the joint occurrences observed in the dataset. The values in these two columns are added in order to check the overall performance of both techniques for each trio.

Estimations based on vine copulas outperformed the results from the traditional method in 65% of the scenarios[36] presented in Table 4.5. Considering each panel separately, we can see that eight trios had better results for the copula approach, i.e. the sum of the last column was smaller than the sum of the fifth column (the exceptions were ABE and BCD). Nevertheless, like in the bivariate comparison (Chapter 3), the vine copula method resulted in higher ratio of underestimation (16.67% against 3.33% from the traditional approach)[37].

We could not find any pattern concerning the performance of the approaches at the different loss levels (from the 4.17% to the 25% highest losses). In some trios (ACE, ADE, BCE, BDE and CDE), the joint losses at the two most extreme levels (4.17% and 8.33%) were better approximated by the trivariate normal method than by the vine copula approach while the other levels presented the opposite results. In other trios (ABC, ABD, and ACD), the predictions from the vine copulas were better for most of the levels (including the two highest ones).

[36] This proportion is very close to the outperformance rate of the copula strategy in the pairwise investigation (63.33%).
[37] The bivariate-copula case had the same underestimation percentage: 16.67%.

Table 4.5 Comparison between estimations of likelihood of joint high default rates (normality vs. copulas estimated according to the entire default's distributions)

Panel A: trio ABC

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00966	0.00027	0.00966	0.00027
8.33%	0.00000	0.03179	0.00197	0.03179	0.00197
12.50%	0.04167	0.05274	0.00595	0.01107	0.03572
16.67%	0.04167	0.08479	0.01273	0.04312	0.02893
20.83%	0.04167	0.09196	0.02266	0.05029	0.01901
25.00%	0.08333	0.11270	0.03597	0.02936	0.04737
Total difference for this trio				0.17529	0.13326

Panel B: trio ABD

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00304	0.00002	0.00304	0.00002
8.33%	0.00000	0.01963	0.00022	0.01963	0.00022
12.50%	0.00000	0.03142	0.00082	0.03142	0.00082
16.67%	0.00000	0.05384	0.00211	0.05384	0.00211
20.83%	0.00000	0.05911	0.00442	0.05911	0.00442
25.00%	0.00000	0.07760	0.00812	0.07760	0.00812
Total difference for this trio				0.24464	0.01572

Panel C: trio ABE

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00000	0.00044	0.00000	0.00044
8.33%	0.00000	0.00001	0.00148	0.00001	0.00148
12.50%	0.00000	0.00106	0.00335	0.00106	0.00335
16.67%	0.00000	0.00385	0.00633	0.00385	0.00633
20.83%	0.00000	0.00564	0.01080	0.00564	0.01080
25.00%	0.00000	0.01391	0.01718	0.01391	0.01718
Total difference for this trio				0.02447	0.03958

Panel D: trio ACD

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00432	0.00015	0.00432	0.00015
8.33%	0.00000	0.02550	0.00145	0.02550	0.00145
12.50%	0.00000	0.03849	0.00524	0.03849	0.00524
16.67%	0.00000	0.06819	0.01256	0.06819	0.01256
20.83%	0.00000	0.08062	0.02407	0.08062	0.02407
25.00%	0.00000	0.10495	0.04004	0.10495	0.04004
Total difference for this trio				0.32206	0.08351

Panel E: trio ACE

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00000	0.00003	0.00000	0.00003
8.33%	0.00000	0.00001	0.00021	0.00001	0.00021
12.50%	0.00000	0.00138	0.00067	0.00138	0.00067
16.67%	0.00000	0.00489	0.00149	0.00489	0.00149
20.83%	0.00000	0.00849	0.00279	0.00849	0.00279
25.00%	0.00000	0.01987	0.00470	0.01987	0.00470
Total difference for this trio				0.03464	0.00990

Panel F: trio ADE

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00000	0.00000	0.00000	0.00000
8.33%	0.00000	0.00005	0.00005	0.00005	0.00005
12.50%	0.00000	0.00240	0.00022	0.00240	0.00022
16.67%	0.00000	0.01079	0.00064	0.01079	0.00064
20.83%	0.00000	0.01559	0.00153	0.01559	0.00153
25.00%	0.00000	0.03809	0.00318	0.03809	0.00318
Total difference for this trio				0.06692	0.00563

Panel G: trio BCD

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00760	0.00003	0.00760	0.00003
8.33%	0.04167	0.03426	0.00023	0.00741	0.04144
12.50%	0.04167	0.06217	0.00076	0.02050	0.04091
16.67%	0.04167	0.07356	0.00184	0.03189	0.03983
20.83%	0.04167	0.08089	0.00373	0.03923	0.03794
25.00%	0.04167	0.09499	0.00674	0.05332	0.03492
Total difference for this trio				0.15996	0.19507

Panel H: trio BCE

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00000	0.00000	0.00000	0.00000
8.33%	0.00000	0.00003	0.00005	0.00003	0.00005
12.50%	0.00000	0.00366	0.00040	0.00366	0.00040
16.67%	0.00000	0.00651	0.00165	0.00651	0.00165
20.83%	0.00000	0.00992	0.00468	0.00992	0.00468
25.00%	0.00000	0.01917	0.01056	0.01917	0.01056
Total difference for this trio				0.03929	0.01734

Panel I: trio BDE

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00000	0.00030	0.00000	0.00030
8.33%	0.00000	0.00005	0.00090	0.00005	0.00090
12.50%	0.00000	0.00423	0.00182	0.00423	0.00182
16.67%	0.00000	0.00741	0.00319	0.00741	0.00319
20.83%	0.00000	0.01004	0.00514	0.01004	0.00514
25.00%	0.00000	0.01930	0.00790	0.01930	0.00790
Total difference for this trio				0.04104	0.01925

Panel J: trio CDE

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00000	0.00001	0.00000	0.00001
8.33%	0.00000	0.00005	0.00012	0.00005	0.00012
12.50%	0.00000	0.00541	0.00071	0.00541	0.00071
16.67%	0.00000	0.00940	0.00236	0.00940	0.00236
20.83%	0.00000	0.01626	0.00582	0.01626	0.00582
25.00%	0.04167	0.02912	0.01182	0.01254	0.02984
Total difference for this trio				0.04366	0.03885

4.5.2 Survival copulas estimated considering the right tails of the default rate distributions

When the survival copulas were estimated with focus on the tails of the default distributions, the superior performance of the vine copula approach increased to 80% of the scenarios (compared to 65% reported in Section 4.5.1)[38]. Apart from the case of BCD, the aggregate absolute difference between the copula estimation and the joint defaults observed in the credit card portfolio for each triplet (last row of the last column in each panel of Table 4.6) was smaller than the absolute difference between the calculations based on the assumption of normality and the ratio of real joint extreme defaults (last row of the fifth column in each panel of Table 4.6). This means that if we need estimations for many risk levels in a particular trio, we are typically better off using the vine copula method as it gives, on average, results closer to the real joint events (see Table 4.6).

[38] This performance was also better than the results from estimations that used right-tail distributions of default rates in the bivariate analysis (70%).

Table 4.6 Comparison between estimations of likelihood of joint high default rates (normality vs. copulas estimated according to tail distributions)

Panel A: trio ABC

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00966	0.00012	0.00966	0.00012
8.33%	0.00000	0.03179	0.00091	0.03179	0.00091
12.50%	0.04167	0.05274	0.00287	0.01107	0.03879
16.67%	0.04167	0.08479	0.00640	0.04312	0.03527
20.83%	0.04167	0.09196	0.01184	0.05029	0.02983
25.00%	0.08333	0.11270	0.01952	0.02936	0.06381
Total difference for this trio				0.17529	0.16873

Panel B: trio ABD

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00304	0.00000	0.00304	0.00000
8.33%	0.00000	0.01963	0.00005	0.01963	0.00005
12.50%	0.00000	0.03142	0.00024	0.03142	0.00024
16.67%	0.00000	0.05384	0.00077	0.05384	0.00077
20.83%	0.00000	0.05911	0.00188	0.05911	0.00188
25.00%	0.00000	0.07760	0.00391	0.07760	0.00391
Total difference for this trio				0.24464	0.00686

Panel C: trio ABE

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00000	0.00000	0.00000	0.00000
8.33%	0.00000	0.00001	0.00000	0.00001	0.00000
12.50%	0.00000	0.00106	0.00001	0.00106	0.00001
16.67%	0.00000	0.00385	0.00003	0.00385	0.00003
20.83%	0.00000	0.00564	0.00008	0.00564	0.00008
25.00%	0.00000	0.01391	0.00024	0.01391	0.00024
Total difference for this trio				0.02447	0.00036

Panel D: trio ACD

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00432	0.00001	0.00432	0.00001
8.33%	0.00000	0.02550	0.00011	0.02550	0.00011
12.50%	0.00000	0.03849	0.00045	0.03849	0.00045
16.67%	0.00000	0.06819	0.00123	0.06819	0.00123
20.83%	0.00000	0.08062	0.00273	0.08062	0.00273
25.00%	0.00000	0.10495	0.00526	0.10495	0.00526
Total difference for this trio				0.32206	0.00980

Panel E: trio ACE

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00000	0.00000	0.00000	0.00000
8.33%	0.00000	0.00001	0.00000	0.00001	0.00000
12.50%	0.00000	0.00138	0.00001	0.00138	0.00001
16.67%	0.00000	0.00489	0.00004	0.00489	0.00004
20.83%	0.00000	0.00849	0.00014	0.00849	0.00014
25.00%	0.00000	0.01987	0.00042	0.01987	0.00042
Total difference for this trio				0.03464	0.00061

Panel F: trio ADE

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00000	0.00000	0.00000	0.00000
8.33%	0.00000	0.00005	0.00001	0.00005	0.00001
12.50%	0.00000	0.00240	0.00007	0.00240	0.00007
16.67%	0.00000	0.01079	0.00026	0.01079	0.00026
20.83%	0.00000	0.01559	0.00072	0.01559	0.00072
25.00%	0.00000	0.03809	0.00168	0.03809	0.00168
Total difference for this trio				0.06692	0.00275

Panel G: trio BCD

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00760	0.00001	0.00760	0.00001
8.33%	0.04167	0.03426	0.00011	0.00741	0.04156
12.50%	0.04167	0.06217	0.00045	0.02050	0.04122
16.67%	0.04167	0.07356	0.00123	0.03189	0.04043
20.83%	0.04167	0.08089	0.00273	0.03923	0.03893
25.00%	0.04167	0.09499	0.00526	0.05332	0.03640
Total difference for this trio				0.15996	0.19855

Panel H: trio BCE

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00000	0.00000	0.00000	0.00000
8.33%	0.00000	0.00003	0.00000	0.00003	0.00000
12.50%	0.00000	0.00366	0.00001	0.00366	0.00001
16.67%	0.00000	0.00651	0.00003	0.00651	0.00003
20.83%	0.00000	0.00992	0.00010	0.00992	0.00010
25.00%	0.00000	0.01917	0.00031	0.01917	0.00031
Total difference for this trio				0.03929	0.00045

Panel I: trio BDE

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00000	0.00000	0.00000	0.00000
8.33%	0.00000	0.00005	0.00001	0.00005	0.00001
12.50%	0.00000	0.00423	0.00005	0.00423	0.00005
16.67%	0.00000	0.00741	0.00018	0.00741	0.00018
20.83%	0.00000	0.01004	0.00052	0.01004	0.00052
25.00%	0.00000	0.01930	0.00125	0.01930	0.00125
Total difference for this trio				0.04104	0.00201

Panel J: trio CDE

Proportion of highest losses	Dataset	Normal	Copula	Difference Normal - Dataset	Difference Copula - Dataset
4.17%	0.00000	0.00000	0.00000	0.00000	0.00000
8.33%	0.00000	0.00005	0.00001	0.00005	0.00001
12.50%	0.00000	0.00541	0.00007	0.00541	0.00007
16.67%	0.00000	0.00940	0.00029	0.00940	0.00029
20.83%	0.00000	0.01626	0.00088	0.01626	0.00088
25.00%	0.04167	0.02912	0.00216	0.01254	0.03950
Total difference for this trio				0.04366	0.04075

Therefore estimations using vine copulas outperform evaluations from traditional models simplified by the assumption of normality and, as shown in the pairwise study (Chapter 3), when the copulas are chosen based exclusively on the upper tail of distributions, the superior performance is even better.

The levels of underestimation were the same found in the previous section: 16.67% and 3.33% for copula-based and normality-based calculations respectively. Note that the underestimation percentage in the trivariate case for the vine model (16.67%) is lower than the underestimation ratio in the bivariate case (26.67%) reported in Chapter 3. This may be evidence that, for datasets spanning "short" periods, copula estimations are less subject to underestimation in higher dimensions (i.e. when vines are employed).

As in Section 4.5.1, we did not identify any specific pattern of the methods' performances across the different levels of losses.

4.6 CONCLUSIONS

Chapter 3 showed that the credit card portfolio analysed in this thesis has a heterogeneous dependence structure (i.e. different best-fit copulas for some pairs of segments). This is likely the case in most loan portfolios given that financial institutions deal with a high number of debtors with distinct characteristics (which tend to generate several types of association). Thus the estimation of joint default rates will be more accurate if the different relationships are incorporated into the calculations.

Vine copulas were employed to investigate the diverse dependence within the credit portfolio because their flexible construction allows the combination of different families of copulas that can give more realistic evaluations of joint default rates.

We consider the case of three segments assessed together and this appears to be the first application of vine structures to credit risk. The dependence of each triplet is characterised by a copula that links the distributions of two segments conditional on a third segment. The linking copula for each triplet was estimated following the two strategies used in Chapter 3: goodness-of-fit tests based on the complete distributions of defaults and only on their upper tail.

The results in both cases revealed several compositions of dependence for triplets of segments that decreased, increased or did not change the probability of simultaneous high losses expressed by each bivariate copula independently. Thus higher-dimension analyses in heterogeneous portfolios may identify potential risk of conjunct extreme default rates not captured in two-dimension studies. The greatest divergence happens when conditional distributions of variables without right-tail dependence are connected by a copula with right-tail association (trio BCE[39] was the best example in this sense).

The same credit card portfolio presented in Chapter 3 was used to compare estimations of joint high default rates in three segments based on assumptions of normality with estimations based on vine copulas. The latter outperformed the former in 65% (80%) of the scenarios tested when the best-fit copulas were inferred according to the default's whole (right tail of) distributions. As in the pairwise analysis, copulas with best fit to upper tails yielded better evaluations than copulas deduced from complete distributions but the vine's outperformance level was higher than the copula's successful rate reported in Chapter 3: 63% and 70% for bivariate copulas derived from the complete distributions of default rates and from their right tails, respectively.

[39] When the copulas were estimated according to the right tail of the default distributions.

The proportion of joint extreme losses underestimated in the vine copula approach was higher than the underestimation from the method assuming normality. However the vine copulas' underestimation level was lower than the underestimation resulted from bivariate copulas. This may indicate that, when short datasets are used, higher-dimension analyses tend to reduce the underestimation of joint losses in copula approaches.

The superiority of the vine approach over calculations based on bivariate copulas (assuming homogeneous dependence in the portfolio) or on the trivariate normal distribution implies that financial institutions may estimate inaccurate probability of joint extreme credit losses (default rates) when using the two last methods. As a consequence, the capital reserved to cover losses according to less precise estimations has higher likelihood of being excessive (which makes banks miss opportunities of investing part of their resources) or insufficient (which menaces the lenders' solvency).

In addition, more accurate evaluations about the risk of simultaneous high losses help the allocation of funds across segments in a more efficient manner such that banks can avoid concentration of loans in segments that tend to be more associated in downturns and do not yield satisfactory return in relation to the risk they represent.

Some further research may expand this pioneering application of vine copulas to credit risk. Longer datasets and other categories of loans apart from credit cards should be used in empirical tests to check the accuracy of vine structures and to confirm whether this approach is better than methods based on normality or bivariate copulas to estimate the likelihood of joint extreme default rates. Canonical vines (as depicted in Figure 4.4) are an alternative to estimate dependence across default rates of many credit segments conditional on a factor that represents the economic scenario.

The literature on credit risk has presented some evidence that the copula approach leads to better evaluation of joint high default rates than methods that presume normality (which is the case of factor credit models). Chapters 3 and 4 have extended such evidence to consumer loans.

In the next two chapters, we use Copula Theory to suggest some improvements to a method derived from factor credit models: the formula used to estimate the capital to be held by financial institutions to cover extreme credit losses (the so-called Basel Accords). Being derived from factor models, Basel method presents the same limitations (as traditional credit risk models) concerning the assumption of normality and the use of linear correlation; copulas are therefore an alternative to reach better results.

CHAPTER 5
COPULA-BASED FORMULAS TO ESTIMATE UNEXPECTED CREDIT LOSSES (THE FUTURE OF BASEL ACCORDS?)[40]

"Discovering consists of looking at the same thing as everyone else does and thinking something different."
(Albert Szent-Györgyi, 1937 Nobel Prize in Physiology and Medicine)

5.1 INTRODUCTION

The current rule to calculate the capital necessary to cover unexpected credit losses (Basel Accords II and III) is based on (single) factor models which define that joint defaults are driven by a latent variable which, in turn, is driven by an unobserved (economic) factor.

The economic factor, the latent variable, and the specific (idiosyncratic) risk for each obligor are assumed to follow the standard normal distribution but there is vast evidence in the literature that those variables are not normally distributed.

The dependence across pairs of latent variables and between each latent variable and the economic factor is measured by the correlation coefficient which is accurate only for normal data and does not detect tail dependence. So, the current model used to calculate the regulatory capital is deficient because it may not identify conjunct extreme occurrences.

To overcome this problem, this chapter proposes the application of copulas to link distributions of latent variables and evaluate unexpected credit losses in financial institutions.

Our contribution is to relax the assumption of normality in the context of capital adequacy in financial institutions and to allow more accurate identification of tail dependence across loan losses (which is particularly important in adverse economic scenarios when credit losses tend to be more associated)[41].

[40] This chapter is a slightly modified version of Moreira (2010).
[41] See references in Chapter 3, Section 3.1.

The latent variables are considered to be survival functions of the probabilities of default (PDs), i.e., high PDs indicate low values of latent variables and vice versa.

While traditional credit risk models use percent values (as cardinal numbers), the copula approach is based on percentiles (ranks) of the variables. Considering that portfolios/segments are taken for homogenous, the expected levels (percentiles) of the latent variables that imply default are equal for all loans. Then, for each pair of debtors, the copula will associate two equal variables (percentiles of latent variables) in extreme conditions and will return the likelihood of both percentiles being simultaneously below a specific level (percentile of the latent variable's historical average in this case). This is equivalent to the probability of potential losses being above the rank of the average (expected) PD.

The implementation of this alternative method is relatively simple and, alike models derived from Merton's approach (Merton, 1974), the model suggested here is based on the interpretation that default happens when the latent variable falls below a cutoff value. The copula method focuses on joint defaults which occur when the latent variables of loans become smaller than their limit percentile at the same time. Losses are unexpected (above the average) when such underlying variables drop even more and reach percentiles smaller than their average's percentile among the values that indicate default. Thus, for a particular level of confidence, "high" unexpected losses will be estimated by a copula that gives the joint probability of the historical latent variable's average being below an extreme percentile.

In principle, a general approach is presented to derive formulas based on any copula found to be representative of loan portfolios. If large datasets on PDs (or default rates) are available, precise models may be built according to the steps proposed in this study.

An example is given for the case where PDs are assumed to be right-tail associated and, consequently, the latent variables present left-tail dependence.

For convenience, the relationship between the latent variables is represented by the Clayton Copula.

Simulations reveal that, in most of the cases, when compared to Basel method, the alternative model yields better estimations of the effective losses in portfolios with tail-dependent probabilities of default (which is expected to be a property of most credit portfolios in the financial market – see some references in Section 5.4).

In around 73% of the scenarios, the copula-based approach outperformed Basel method for at least one of the three credit classes analysed (revolving consumer, mortgage, and “other retail”). On average, the new method was better for all three categories in 52% of the cases. The results were sensitive to the confidence specified and the shape of the loss distribution. Normally-distributed losses generated the worst estimations for the suggested model at the confidence level used while the other three distributions tested (Exponential, Beta, and Gamma) resulted in an outperformance ratio of 75%.

This chapter is organised as follows. Basel Accords are addressed in the next section. Then Copula Theory is discussed. Section 5.4 summarises a general approach to derive formulas based on assumed or empirically found dependence between probabilities of default. In Section 5.5, PDs are presumed to be right-tail dependent (i.e. high losses are more associated) and a formula based on the Clayton Copula is derived to estimate unexpected losses. Next, the results from the formula presented in the prior section are compared to the capital calculated by the Basel formula. Section 5.6 concludes.

5.2 BASEL ACCORDS

The Basel Accord from 1988 stipulated that the capital charge on assets was 8% of the risk weighted assets. But due to many drawbacks in this Accord (see De Servigny and Renault, 2004), new rules were issued in June 2004 (Basel II).

The Basel II Accord was based on three “pillars”: minimum capital requirements, Supervisory Review, and market discipline. Banks were allowed to use Internal Ratings Based approaches (IRB) to calculate the capital required and to do so, institutions should group their assets into homogenous “buckets” (segments, classes) with respect to their credit quality (see BCBS, 2006).

A new reform was proposed at the end of 2009 (Basel III) but it did not alter the formulas presented in Basel II for calculating extreme credit losses[42]. Since this calculation is the main purpose here, we will refer to Basel II as the current method to estimate the capital necessary to cover unexpected credit losses.

Although the method adopted in Basel II (and kept in Basel III) improved the calculations established in the first Basel Accord, the new approach also has some limitations. It assumes normally-distributed loans’ performance and uses the correlation coefficient that does not capture oscillations in dependence when the level of variables changes. Thus, this may lead to excessive capital required in good economic scenarios or scarce requirements in downturns.

For each segment, the capital required to cover unexpected losses in credit portfolios is calculated as the unexpected losses adjusted by the portfolio maturity. In mathematical terms:

$$[LGD * K_V - LGD * PD] * Maturity = [LGD * (K_V - PD)] * Maturity \qquad [\ 5.1\ ]$$

where LGD is the “loss given default”, i.e. the percentage of exposure the lender will lose if borrowers default and PD stands for probability of default. $Maturity$ corresponds to the maturity of *corporate* loans (i.e., not applied to consumer debt) and is added to the calculation in order to give higher weight to long-term credits which are known to be riskier. It is calculated as:

$$maturity = \frac{1 + (M - 2.5) * b(PD)}{1 - 1.5 * b(PD)}$$

[42] Information retrieved at http://www.bis.org/bcbs/basel3.htm on November 16th, 2010.

where $b(PD) = (0.11852 - 0.5478 * \log(PD))^2$ and M is the average maturity of the credit portfolio.

The other term in [5.1], K_V, is the expected default rate at the 99.9% percentile of the PD distribution ("Vasicek Formula") - see Vasicek (1991, 2002). In general, K_V follows the main presumptions of factor models (see, e.g., Gordy, 2003) where each latent variable (y_i) is a linear function of an unobserved single factor (systematic risk, E) and specific characteristics of the respective obligor (idiosyncratic risk, ε_i). The latent variable, the single factor and the idiosyncratic risk terms are assumed to be standard normally distributed. The economic factor impacts all obligors equally (same correlation $\sqrt{\rho}$) and the latent variables are considered equicorrelated (same ρ for all pairs). This leads to expression [2.4] presented in Chapter 2, Section 2.4, and restated below for convenience:

$$y_i = E\sqrt{\rho} + \varepsilon_i \sqrt{1-\rho} \qquad [5.2]$$

For each loan i, the probability of default is the likelihood that the latent variable y_i becomes smaller than the cutoff y_c, that is, $PD = \Pr[y_i < y_c]$. Extreme credit losses happen when the economy E reaches an extremely unfavourable level $e*$. In other words, these high losses are the probability of default conditional on poor economic status. Representing this probability as $PD*$, we have $PD* = \Pr[y_i < y_c \mid E = e*]$ and using [5.2]:

$$PD* = \Pr[E\sqrt{\rho} + \varepsilon_i \sqrt{1-\rho} < y_c \mid E = e*]$$

Solving for ε_i and replacing E with $e*$:

$$PD^* = \Pr\left[\varepsilon_i < \frac{y_c - e^*\sqrt{\rho}}{\sqrt{1-\rho}}\right]$$

As mentioned in Chapter 2, Section 2.4, ε_i is presumed to be normally distributed with mean 0 and variance 1. Thus, the previous equation turns into:

$$PD^* = \Phi\left(\frac{y_c - e^*\sqrt{\rho}}{\sqrt{1-\rho}}\right)$$

where Φ indicates the *cdf* of the standard normal distribution.

Since y_i is also normally distributed, $PD = \Phi(y_c)$ which implies that $y_c = \Phi^{-1}(PD)$, i.e. the cutoff of the latent variable below which default occurs is the inverse of the normal distribution, Φ^{-1}, evaluated at PD. Basel II demands confidence of 99.9% which means that the capital is supposed to be sufficient to cover the losses whenever the economy is above (better than) the 1st percentile of its distribution (also assumed to be normal). Hence the extreme adverse scenario e^* is given by $\Phi^{-1}(0.001)$. Due to two properties of the standard normal distribution (symmetry and mean 0), $\Phi^{-1}(0.001) = -\Phi^{-1}(0.999)$. Using this fact and replacing e^* with $-\Phi^{-1}(0.999)$ and y_c with $\Phi^{-1}(PD)$ in the prior equation, we get the formula presented in Basel II (here the extreme loss, PD^*, is denoted as K_V):

$$K_V = \Phi\left(\frac{\Phi^{-1}(PD) + \sqrt{\rho}\Phi^{-1}(0.999)}{\sqrt{1-\rho}}\right) \qquad [\ 5.3\ ]$$

where:

Φ and Φ^{-1} represent the standard normal cumulative distribution and its inverse, respectively;

PD, as before, is the probability of default of the loan portfolio (average);

$\Phi^{-1}(PD)$ is used to derive the default threshold (i.e. the cutoff level of obligors' assets below which default occurs);

$\Phi^{-1}(0.999)$, which is equal to $-\Phi^{-1}(0.001)$, is the level of the economy chosen to represent an extreme scenario in which unexpected losses may occur. Therefore, the systematic factor is assumed to be normally distributed. Intuitively, this is the confidence level (99.9%) for the default rate; and

Rho (ρ) is the correlation between returns of obligors' assets. $\sqrt{\rho}$ is the linear correlation between the unobserved systematic factor and those asset returns. In Basel II, the correlation between asset returns is calculated as a function of PD and (in the case of corporate debt) the size of debtors (measured in terms of annual sales). For the sake of brevity, the formula and parameters used to estimate ρ will not be presented here. See BCBS (2005, 2006) for more details.

Readers interested in more details about the derivation of the Vasicek Formula (K_V) should consult, for instance, Schönbucher (2000), Perli and Nayda (2004), and Crook and Bellotti (2010). For more information about the main risk parameters used in Basel II (including PD and LGD), see Engelmann and Rauhmeier (2006).

In summary, $LGD*K_V$ in expression [5.1] gives the total potential loss and $LGD*PD$ represents the expected losses. The difference between them is therefore the unexpected losses. The proposed formulas in this study are limited to replace the term $(K_V - PD)$, which expresses the unexpected default rate, and do not consider possible shortcomings in the computation of LGD and the maturity adjustment.

Some models have been proposed to transform [5.3] into another expression that does not have the limitation regarding the assumption of normality. Departing from [5.2], Hull and White (2004) relax the distributions of y_i, E and

ε_i[43], such that they can, for example, present heavy tails (which tends to increase the joint occurrences of extreme realisations of the latent variables). Representing the distributions of those three variables respectively by F, G and H and following the same steps that derived [5.3] from [5.2], the expression to estimate the probability of default conditional on an unfavourable economic status (the worst 0.1% scenario, i.e. with confidence of 99.9%) turns into:

$$\mathrm{Pr}ob[y < y_c \mid E = e^*] = H\left(\frac{F^{-1}(PD) - \sqrt{\rho}G^{-1}(0.001)}{\sqrt{1-\rho}}\right)$$

where e^* indicates an extreme adverse economic scenario and can be calculated as the inverse distribution of E evaluated at 0.001 (since the critical level was set at 0.1%). PD is the historical probability of default and ρ is the linear correlation between returns of obligors' assets. Obviously, the expression above cannot be solved unless the shapes of the three distributions F, G and H are known.

Some studies, such as Bluhm et al. (2002), Kostadinov (2005) and Kang (2005), have suggested the Student t distribution for E and ε_i to characterise the existence of more events (than the normal distribution) in the tails. In this case, it is not possible to define the distribution of the latent variable in [5.2] and the probability of default in downturns (at the 0.1% worst scenario) is:

$$\mathrm{Pr}ob[y < y_c \mid E = e^*] = T_v\left(\frac{F^{-1}(PD) - \sqrt{\rho}T_v^{-1}(0.001)}{\sqrt{1-\rho}}\right)$$

where T_v is the Student t distribution with v degrees of freedom. Given that the latent variable's distribution F remains unknown, the preceding likelihood

[43] Provided that they are scaled with mean zero and variance one.

cannot be calculated. In view of the impossibility of the estimation of the probability of default in adverse economic scenarios when one (or more) of the variables in [5.2] are not normally distributed, we propose a different setup to incorporate Copula Theory into this analysis and capture potential tail dependence even if we do not know *any* of the distributions of the latent variable, the economic factor and the idiosyncratic factors (which is the reality in financial institutions).

5.3 USING COPULAS TO ESTIMATE UNEXPECTED CREDIT LOSSES: A GENERAL APPROACH

5.3.1 Characterisation of default in the copula approach

Traditional approaches employed in the financial industry, such as CreditMetrics® and KMV®, incorporate the basic idea of structural models and assume that default happens when a latent variable (for example, the log-return of debtors' assets) falls below a cutoff point (the outstanding debt). The latent variable (Y) follows a stochastic process[44] and, at any time, the possible values of Y are lognormally distributed. Figure 5.1 shows that, at time t for example, the mean of this distribution is the expected value of Y and the asset return (latent variable) is equal to y_t. The probability of default (PD) is given by the area below the cutoff (denoted y_c in Figure 5.1) under the curve of the Y's distribution. In other words, PD is the probability of the latent variable (Y) being smaller than that particular value (the threshold y_c which is a theoretical value that represents the outstanding debt) at time t. In practice, this characterises situations where the obligors' assets become smaller than the value borrowed.

[44] See Merton (1974) for details on this stochastic process.

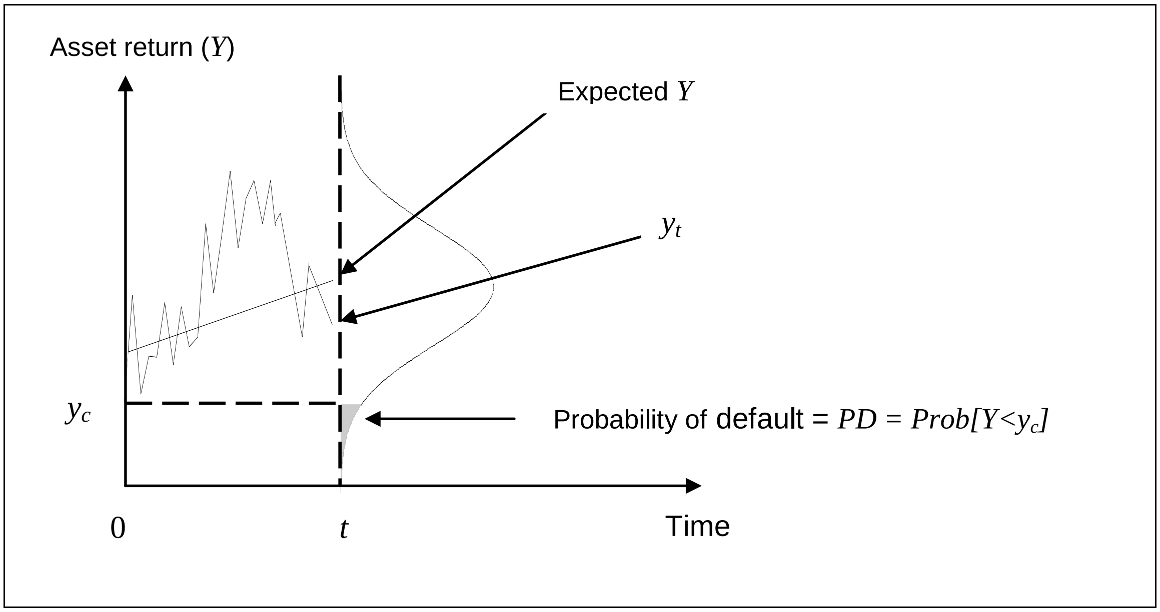

FIGURE 5.1 (based on Crouhy et al., p. 74)
Diagram representing default in structural models as the probability of a latent variable being below a cutoff value (area below y_c at time t).

In this copula-based method, PD is viewed in a portfolio perspective and is defined as the area below a cutoff in the *joint* distribution of latent variables relative to the loans that compose the portfolio. Here, such latent variables are supposed to have a symmetrically inverse relationship with the probability of default. This means that when the latent variable decreases (increases), PD increases (decreases) at the same "degree". The identical magnitude ("degree") of the variables' opposite movements will be expressed by their percentiles in their respective distributions. Hence high (low) levels of PDs are associated with low (high) levels of the latent variables and when PD moves p percentiles in its distribution, Y moves p percentiles in its respective distribution in the opposite direction.

This symmetric inverse behaviour may be captured by representing each latent variable (Y) as a survival function[45] of PD, which implies that the cumulative

[45] The use of the subscribed "t" to indicate the time dependence in survival functions was relaxed.

distribution of the latent variable is equal to one minus the cumulative distribution of the associated PD:

$$F_Y(y) = \overline{F}_{PD}(pd) = 1 - F_{PD}(pd) \qquad [\ 5.4\ ]$$

Note that the cumulative distributions of Y and PD, $F_Y(y)$ and $F_{PD}(pd)$ respectively, give the percentiles represented by the points y and pd. For instance, $F_Y(y) = 0.20$ indicates that y is the 20th percentile of F_Y. Thus, if a specific value of the latent variable y is the p^{th} percentile in F_Y, the PD related to that value of y will be the $1 - p^{th}$ percentile in F_{PD}.

According to [5.4], Y may be interpreted as the probability of non-default and expresses the "quality" of debtors. This idea resembles the survival function used by Li (2000) to define the likelihood that a security will reach a specific age. The higher this probability, the higher the asset quality.

Since we are using cumulative distributions of the variables Y and PD, $F_Y(y)$ and $F_{PD}(pd)$ respectively, Copula Theory may be applied and the resultant calculations are suitable for any kind of loss distribution.

As in the Basel approach, the capital needed to cover unexpected losses will be separately determined for each segment considered homogeneous in terms of credit quality. This means that PD is presumed identical for all loans in each segment and $F_{PD}(pd)$ values are also equal. Therefore, the average (expected) Y is the same for every debtor within the segment and so is $F_Y(y)$. This is true regardless of the number of debtors.

The estimation of unexpected losses depends on an average point considering only the occurrences below the latent variable's cutoff. Figure 5.2 – Panel A shows contours of the joint cumulative distribution of the latent variables and represents the distinction between expected and unexpected losses in this context.

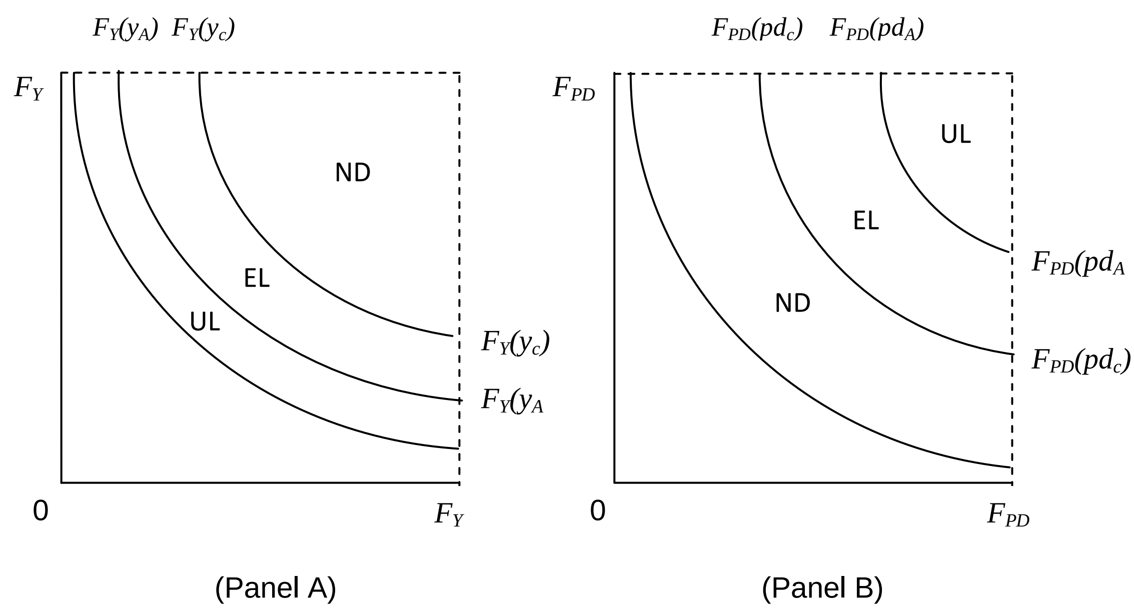

FIGURE 5.2 – Contour plots of cumulative distributions representing expected losses (EL), unexpected losses (UL), and non-default (ND) in a copula context (for homogenous portfolios, where the relationship between any two loans represent all the associations in the pool). In Panel A, default at the portfolio level happens if the cumulative distributions (*cdfs*) of the latent variable, $F_Y(y)$, of both loans fall below a specific point, $F_Y(y_c)$; when each $F_Y(y)$ is smaller than $F_Y(y_c)$ but greater than $F_Y(y_A)$, the losses are expected; and when those *cdfs* drop below the point that indicates average default, $F_Y(y_A)$, the losses are unexpected. Panel B shows ND, EL, and UL under the perspective of the probability of default which are equal to the equivalent areas in Panel A. $F_{PD}(pd_c) = 1 - F_Y(y_c)$ and $F_{PD}(pd_A) = 1 - F_Y(y_A)$ are, respectively, the *cdfs* of PD above which default and unexpected losses happen. Since the focus is on cumulative distributions, both panels are valid regardless of the PD distribution's family. The shapes will depend on the copulas used.

The shapes of the level curves are illustrative and will vary according to the copula used. Along each axis we represent values of F_Y for each of the two loans considered. When the cumulative distribution (*cdf*) of each latent variable Y, $F_Y(y)$, in a homogeneous portfolio falls below the cumulative distribution of the cutoff $F_Y(y_c)$, all obligors default at the same time. The losses are considered expected (area EL in Figure 5.2 – Panel A) while each $F_Y(y)$ keeps falling from $F_Y(y_c)$ until $F_Y(y_A)$, the latter representing the cumulative distribution of the latent variable Y evaluated at the average of its historical values in scenarios of default. On the other hand, if the latent variable becomes smaller than that average (area UL in Figure 5.2 – Panel A), the losses are unexpected (meaning that the obligors' conditions – whose proxy is the latent variable – got worse than usual).

Panel B of Figure 5.2 displays the contours referent to the "default status" of a pair of loans (which includes non-default and default statuses). The axes represent the cumulative distributions of the default statuses, F_{PD} (one for each loan). When the default status for both loans is smaller than a specific point (denoted pd_c), these distributions are simultaneously below $F_{PD}(pd_c)$ and there is no joint default (area ND). When both F_{PD} are greater than the threshold $F_{PD}(pd_c)$ but smaller than the point $F_{PD}(pd_A)$, the probability of default is at the expected level (area EL). When the default statuses of both loans go beyond pd_A the concurrent losses become higher than the historical average (expected losses) and are therefore unexpected (area UL).

We can relate the F_Y distributions (shown in Panel A) to F_{PD} distributions (shown in Panel B). Given that the latent variables and the losses for each loan have inversely symmetric *cdfs*, $F_Y(y) = 1 - F_{PD}(pd)$, the joint function $H(y_A, y_A)$ is equivalent to the area above the *cdfs* of the average probability of default in a complete PD distribution ($F_{PD}(pd_A)$ in Figure 5.2 – panel B), i.e. a distribution that includes non-default status. Thus, both areas UL in Panels A and B of Figure 5.2 are equal to each other and indicate unexpected losses[46].

What we should estimate is the likelihood of the joint probability of default for two obligors being above its average, i.e. the probability that both $F_{PD}(pd)$ are greater than the average $F_{PD}(pd_A)$. Recalling the concept of Survival Copulas in Chapter 2, Section 2.3.1, and that each debtor has the same PD, we have:

[46] ND and EL are also equal in both panels but they do not affect the calculations here (which pertain to the unexpected losses).

$$\overline{H}(pd_A, pd_A) = \hat{C}(\overline{F}_{PD}(pd_A), \overline{F}_{PD}(pd_A)) = \hat{C}(1 - F_{PD}(pd_A), 1 - F_{PD}(pd_A)) \qquad [\,5.5\,]$$

In the prior formula, pd_A stands for the average of the historical probability of default and the notation $\overline{H}$ refers to a joint *survival* function. The expression gives the probability of both PDs being above the historical average pd_A at the same time. The Survival Copula $\hat{C}$ links the two univariate survival functions of $\overline{F}_{PD}(pd_A) = 1 - F_{PD}(pd_A)$ to the bivariate function.

Now, applying the definition introduced in [5.4] to the average PD, we have $F_Y(y_A) = 1 - F_{PD}(pd_A)$ and [5.5] becomes:

$$\overline{H}(pd_A, pd_A) = \hat{C}(F_Y(y_A), F_Y(y_A)) = H(y_A, y_A) \qquad [\,5.6\,]$$

where pd_A and y_A are the historical average of PD and of the latent variable, respectively; for homogenous pools of borrowers i and j, such that $PD_i = PD_j = PD$ and $Y_i = Y_j = Y$, $\hat{C}$ is a copula that returns $\Pr(PD > pd_A, PD > pd_A)$ = $\Pr(Y < y_A, Y < y_A)$ the probability of both PDs (latent variables) being above (below) their observed average up to the moment or, in other words, the probability of unexpected losses.

5.3.2 Finding the ranks of the latent variable

To apply this copula model we need the *whole* distribution of the latent variable so that we can calculate the cumulative distribution function (*cdf*) of Y associated with the point of historical average loss, $F_Y(y_A)$. Given a group of obligors in default, the *cdf* of the cutoff $F_Y(y_c)$ would be obviously 1 and any area calculated under this circumstance, would return the likelihood of PD being below or above a point and not the PD itself.

Based on Figure 5.2 – Panel A, we see that to find the unexpected losses (UL) we need to know the *cdf* $F_Y(y_A)$ in the *complete* distribution of Y (i.e. including non-default status, ND). However, in principle, we do not have enough information to find that value. Otherwise, we could use $F_Y(y_c)$ to calculate the total losses (EL + UL) and subtract EL, which is known (the average PD of the portfolio). But, again, we cannot find $F_Y(y_c)$ using solely the information available so far.

One way to start solving this problem is considering a relationship between $F_Y(y_A)$ and $F_Y(y_c)$. Figure 5.3 – Panel A illustrates the density $f_{default}(.) = f(. \mid Y < y_c)$ of a latent variable Y that includes only default cases (i.e. all observations have values below the cutoff that indicates default). $F_{default}$, defined as $F_{default}(.) = F_Y(. \mid Y < y_c)$, is the correspondent cumulative distribution and $F_{default}(y_A) = F_Y(y_A \mid Y < y_c)$ is the *cdf* of Y evaluated at the average latent variable in that distribution limited to default cases. The latter can be estimated from datasets of PDs taken over several periods by finding the *cdf* of PD evaluated at the average PD and applying [5.4]: $F_{default}(y_A) = 1 - F_{default}(pd_A)$. The cutoff y_c is the largest value in that density function, so $F_{default}(y_c) = 1$.

The distribution of the latent variable becomes complete if we add the non-default cases (when the latent variable is higher than the cutoff value) as in Figure 5.3 – Panel B. The complete distribution F_Y is not observable and may have any shape.

As an example, consider that the latent variables represent borrowers' asset returns. For the sake of simplicity, the debt of all obligors will be assumed equal but this presumption can be easily relaxed if we work with the percentage of asset returns over the (different) liability of each obligor. If all debts are equal to 100 monetary units, y_c = 100 and default happens when $Y <$ 100. So, debtors may default with different values of asset returns below the theoretical cutoff y_c. Assume, for example, that a portfolio has five defaulters with y = {35, 75, 80, 90, 95}. In this case, the average asset return (y_A) among the debtors who failed their repayments is 75 monetary units. In Panel B, the obligors that compose the areas UL (unexpected losses), EL (expected losses) and ND (non-default) have Y < 75, 75 ≤ Y < 100, and Y ≥ 100, respectively.

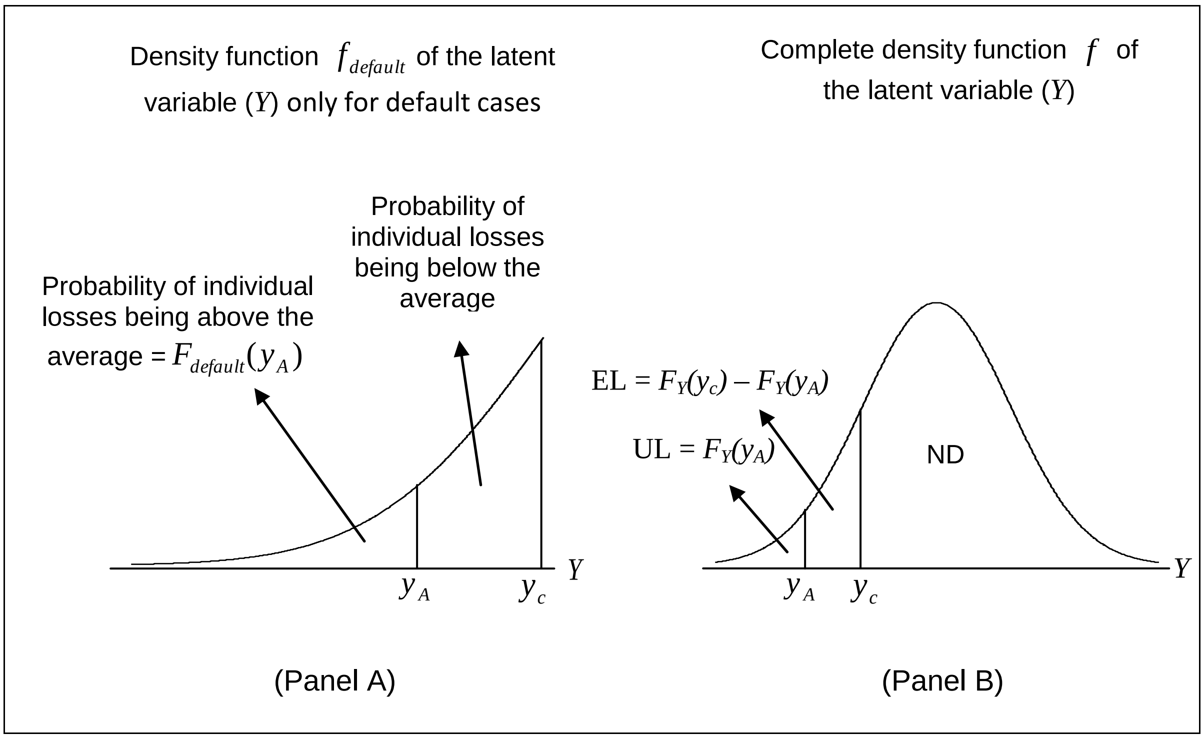

FIGURE 5.3 – Density function of the latent variable (Y). The shapes are merely illustrative. Panel A displays only the cases where losses happened whilst Panel B includes levels of Y that did not result in default (above the cutoff y_c). The *cdf* of Y at y_A, $F_{default}(y_A)$, is equal to one minus the *cdf* of PD assessed at the average PD, which can be inferred from datasets. $F_{default}(y_c) = 1$. UL, EL, and ND represent unexpected losses, expected losses, and non-default, in that order. $F_Y(y_A)$ and $F_Y(y_c)$ are, respectively, the *cdfs* of the latent variable related to the historical average losses and to the cutoff value below which defaults happen.

We cannot observe Y or its distribution but we know that its *cdf* is equal to one minus the *cdf* of the associated PD in the PD distribution. So, if expected losses (average PD = pd_A) are, for instance, 5% and this value is the 60th percentile in its distribution restricted to default cases, according to [5.4], the *cdf* of y_A in the latent variable's distribution, $F_{default}(y_A)$, will be $1 - F_{default}(pd_A) =$. $1 - 0.60 = 0.40$. This reasoning also works for the complete (unobservable) distribution F_Y.

Regardless of the size of the non-default area (ND) in Figure 5.3 – Panel B, y_A and y_c are the same in both distributions (F_Y and $F_{default}$) and their *cdfs* indicate the proportion of data occurrences below those specific points. $F_{default}(y_A)$, for instance, gives the number of Y observations in distribution $F_{default}$ below y_A, $n_A^{default}$, divided by the total observations, $n^{default}$. Thus:

$$\frac{F_{default}(y_A)}{F_{default}(y_c)} = \frac{n_A^{default} / n^{default}}{n_c^{default} / n^{default}} = \frac{n_A^{default}}{n_c^{default}}.$$

Regarding the distribution F_Y, let n_A, n_c, and n denote, respectively, the number of observations below y_A, below y_c, and in the complete distribution. Following the reasoning in the prior paragraph: $\frac{F_Y(y_A)}{F_Y(y_c)} = \frac{n_A / n}{n_c / n} = \frac{n_A}{n_c}$.

Since no data is included below y_c when the non-default area (ND) is added to $F_{default}$ in order to generate the entire distribution F_Y, $n_A^{default} = n_A$ and $n_c^{default} = n_c$.

Therefore, $\frac{F_{default}(y_A)}{F_{default}(y_c)} = \frac{F_Y(y_A)}{F_Y(y_c)}$. As stated before, $F_{default}(y_c) = 1$, thus $F_Y(y_c)$ is always equal to $F_Y(y_A) / F_{default}(y_A)$.

From Figure 5.2 – Panel A that represents homogenous segments/portfolios (same PD for all loans), it is easy to see that the joint area below y_c minus the joint area below y_A is equal to the expected probability of default (EL). In copula terms,

$$\hat{C}(F_Y(y_c), F_Y(y_c)) - \hat{C}(F_Y(y_A), F_Y(y_A)) = \qquad [5.7]$$
$$\hat{C}(F_Y(y_A)/F_{default}(y_A), F_Y(y_A)/F_{default}(y_A)) - \hat{C}(F_Y(y_A), F_Y(y_A)) = PD$$

where $F_Y(y_A)$, the *cdf* of the latent variable evaluated at its historical average, is the only unknown variable, $F_Y(y_c) = F_Y(y_A)/F_{default}(y_A)$, $F_{default}(y_A)$ is the *cdf* of Y at its historical average in the distribution restricted to $Y < y_c$, and PD expresses the expected (average) probability of default (EL). The notation $\hat{C}$ (from [5.6] and based on Nelsen, 2006) was kept to indicate that we are dealing with a survival copula from a PD standpoint[47]. The existence of a closed-form solution to calculate $F_Y(y_A)$ will depend on the copula chosen or empirically found to represent the association between the latent variables of the loans.

After $F_Y(y_A)$ is estimated, the joint distribution $H(y_A, y_A)$ may be calculated as the copula $\hat{C}(F_Y(y_A), F_Y(y_A))$ and will express the mean unexpected losses in a particular period (the sum of percent losses above the average in a period divided by the number of unit times considered – months, for instance).

However, in bank regulation, the major concern is the maximum potential loss. In this copula-based method, the risk of severe unexpected losses comes from possible variations in the *cdf* of the latent variable evaluated at its past average (= expected latent variable), i.e. changes in $F_Y(y_A)$ that may reach extreme values while y_A is kept constant (the historical average). The augment of that *cdf* is interpreted as a response to the deterioration of the economic status.

---

[47] Since $F_Y(.) = 1 - F_{PD}(.) = \overline{F}_{PD}(.)$, [5.7] corresponds to:
$\hat{C}(\overline{F}_{PD}(pd_c), \overline{F}_{PD}(pd_c)) - \hat{C}(\overline{F}_{PD}(pd_A), \overline{F}_{PD}(pd_A)) = PD$.

This situation can be depicted with the support of Figure 5.3 – Panel B. In downturns, latent variables smaller than y_A tend to appear more frequently. In these circumstances, the percentage of non-default (ND) drops and, as the expected losses (EL) stay unaltered (it is set as the historical average), the unexpected losses (UL) rise. Therefore the ratio UL/EL goes up and so does $F_Y(y_A)$. It is worth noting that y_A remains steady and each new $Y < y_A$ makes y_A "move" to the right side *at the density representation* and get closer to y_c. The risk is "how far" y_A can go, i.e. how close to $F_Y(y_c)$ $F_Y(y_A)$ can get. In the example given above (pertaining to the five-defaulter portfolio with y = {35, 75, 80, 90, 95} and y_A = 75 *before the downturn*), new defaulters in an unfavourable economic scenario would likely have Y < 75 (latent variables smaller than the historical average) which would make $F_Y(y_A)$ increase.

In order to estimate this potential increment of $F_Y(y_A)$, we should find an extreme *cdf* of the average latent variable in the distribution resultant from the inclusion of smaller latent variables, $F_{EXT}(y_A) = F_Y(y_A \mid Y < y_c$ *and Economic Status* = *downturn*), as illustrated in Figure 5.4 that follows the intuition of Figure 5.3. In Panel A, the area below the extreme percentile of each loan, $F_{default\_EXT}(y_A)$, is the model confidence, i.e., the probability that $Y < y_A$ in $f_{default\_EXT}$. The *cdf* of Y at the cutoff value of the latent variable below which default occurs is equal to one, $F_{default\_EXT}(y_c) = 1$.

In Panel B, values of Y above y_c (non-default status) are included in the density function. UL, EL, and ND stand for unexpected losses, expected losses, and non-default, respectively. UL is given by $F_{EXT}(y_A)$ which is associated with $F_{EXT}(y_c)$ as $F_{EXT}(y_c) = F_{EXT}(y_A) / F_{default\_EXT}(y_A) = F_{EXT}(y_A) / confidence$.

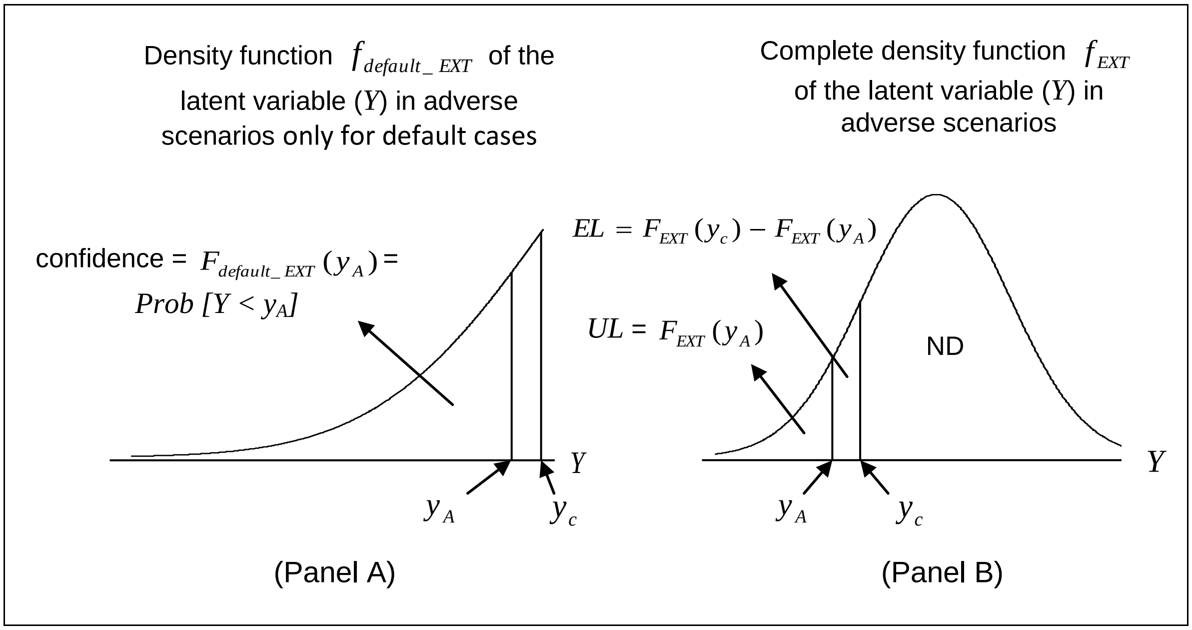

FIGURE 5.4 – Density function of the latent variable (Y) depicting a situation where the average of the historical latent variable (y_A) becomes an extreme percentile in an unfavourable scenario. The shapes are merely illustrative.

The confidence required will express the ratio UL/(UL+EL) and may be understood as a measure of the economy's degradation. $F_{EXT}(y_A)$ will be a proportion of $F_{EXT}(y_c)$ such that $F_{EXT}(y_c) = F_{EXT}(y_A) * confidence$, with confidence between 0 and 1. In this fashion, when confidence equals 100%, all losses are unexpected. Conversely, when it approaches zero (upturns), small unexpected losses are supposed to happen. So, like in Basel II and factor models, the latent variables of loans are driven by the (unobserved) economic status. Here, the latter is captured by the variation of the former which, in turn, is inferred from available data on probabilities of default.

Using the example mentioned earlier, in which the latent variable is interpreted as obligors' asset returns, y_A is still 75 (the historical average) but due to the severe economic conditions, asset returns lower than the average (Y < 75) are included in the distribution and y_A represents a higher percentile in the new distribution F_{EXT} that includes latent variables observed in downturns, i.e. $F_{EXT}(y_A) > F_Y(y_A)$.

The copula calculated for the extreme percentiles of y_A will give the maximum unexpected loss with the confidence demanded which defines the location of the average latent variable in the new distribution F_{EXT}. Following the same reasoning in [5.7], we can find the extreme *cdf* for each loan, $F_{EXT}(y_A)$, doing:

$$\hat{C}(F_{EXT}(y_A)/confidence, F_{EXT}(y_A)/confidence) - \hat{C}(F_{EXT}(y_A), F_{EXT}(y_A)) = PD \quad [5.8]$$

PD is the average probability of default (EL), $F_{EXT}(y_c) = F_{EXT}(y_A)/confidence$ and $confidence \in (0,1]$ establishes the *cdf* of the average latent variable for each obligor in an adverse economic scenario.

The final formula is intended to replace the term $(K_V - PD)$ in [5.1]. Thus, the capital to cover unexpected losses will be:

$$[LGD * \hat{C}(F_{EXT}(y_A), F_{EXT}(y_A))] * Maturity$$

5.3.3 Defining the copula to be used

If large datasets on probabilities of default are available, the dependence across pairs of latent variables may be found through the estimation of the best copula for $"1-PD"$. Therefore it is not necessary to estimate the best copula that expresses the dependence between PDs. What matters is the copula that will represent the dependence across the latent variables (which can be interpreted as returns of debtors' assets or "time until default", for instance). To estimate such dependence it suffices to have a series of PDs from a "homogeneous" credit segment/portfolio.

Durrleman et al. (2000) and Cherubini *et al*. (2004, Chapter 5), for example, present some methods that can be used to empirically find the parameter, for each copula family, with the best fit to a dataset (see also Chapter 3, Section 3.2.1). A practical way to find the copula's parameter is to estimate it from the kendall's tau of $1-PD$ (by using [2.1] presented in Chapter 2) which is the same kendall's tau for PDs (which are observable)[48].

Berg (2009) and Genest *et al*. (2009) describe some goodness-of-fit tests that allow us to decide which copula (considering the estimated parameters) gives the best expression of the dependence related to the variables analysed (see also Chapter 3, Section 3.2.2).

The use of empirically-found copulas gives more realistic results than the use of assumed copulas because the probability of unexpected losses and the dependence between the variables come from "real" data (PDs).

The following example shows the application of the model if we assume that high PDs are more linked than low PDs.

5.4 MODEL APPLICATION: AN EXAMPLE FOR RIGHT-TAIL-DEPENDENT LOSSES

5.4.1 Assumptions

As stated in Chapter 3, Section 3.1, many studies have shown that asset returns present stronger association when they (returns) are at lower levels and this conclusion has been extended to credit assets. Based on this, it is assumed in this section that PDs (probabilities of default, credit losses) have upper tail dependence (which means that high PDs are more correlated than the other levels or, in other words, large losses of different obligors tend to be more associated whereas small losses are not very linked). This relationship can be represented by copulas such as Gumbel, Joe, Galambos, and Hüsler-Reiss. The Gumbel was chosen because, among those copulas cited, it has been more studied and its properties are better known.

---

[48] Alternatively, [2.2] can be used to estimate the copula parameter as a function of Spearman's rho (ρ_S). In the simulations run for this study, the results based on τ and ρ_S usually matched up to the second decimal place.

The scatter plot of a Gumbel-dependent random variable X $(0 \leq X \leq 1)$ looks like Figure 5.5. Consequently, the plot of the symmetrically inverse variable $1 - X$ will be like Figure 5.6.

Those figures are suitable for representing the dependence between *PDs* of loans and between their latent variable Y respectively, such that $F_Y(y) = 1 - F_{PD}(pd)$ as defined in [5.4]. The Clayton Copula is a good representation for the second type of dependence (between latent variables) that indicates lower tail association. This relationship could be expressed by other copulas that express lower-tail dependence (Raftery, for instance) but the Clayton Copula was chosen because it has been more studied and its formula is more tractable than the other alternatives.

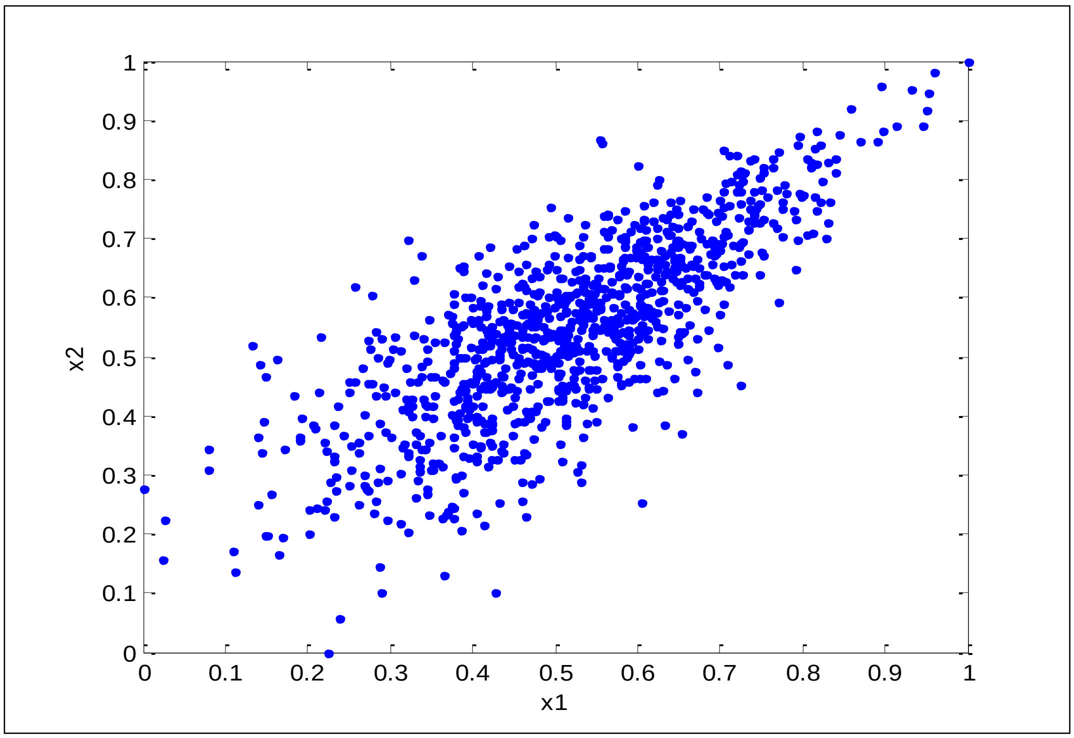

FIGURE 5.5 – Two random variables with Gumbel dependence (upper-tail dependence).

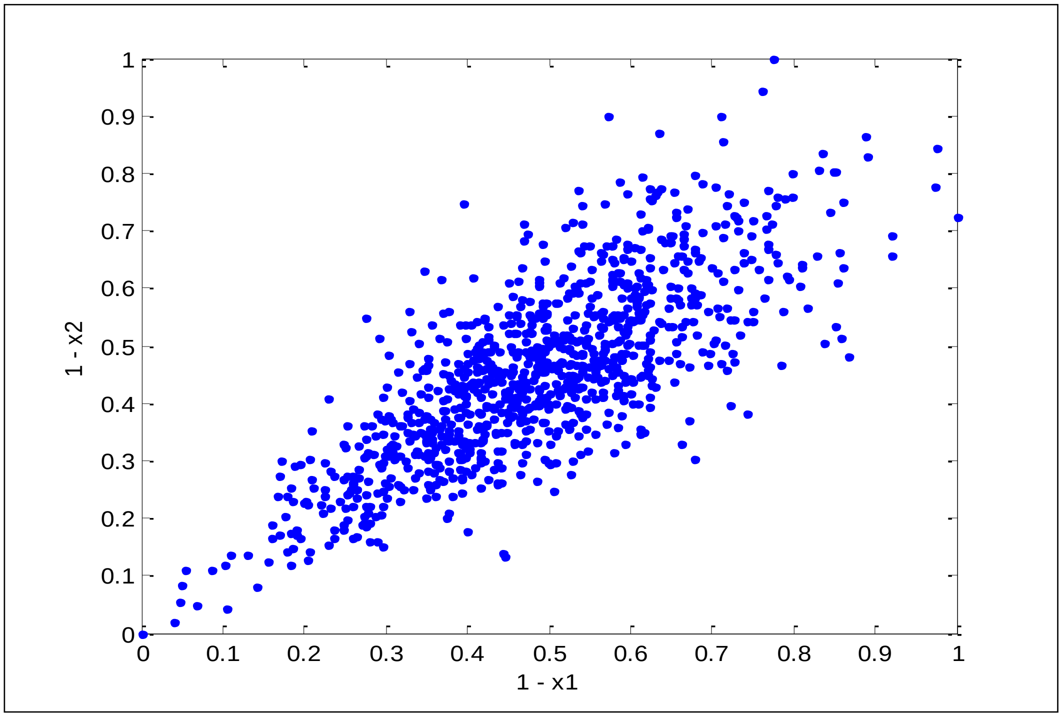

FIGURE 5.6 – Two random variables with lower-tail dependence.

5.4.2 The formula

We are interested in calculating the joint probability of the latent variable's historical average being below the percentile of an extreme point that indicates joint unexpected losses in adverse scenarios. To do so, we should estimate the copula $\hat{C}(F_{EXT}(y_A), F_{EXT}(y_A))$ where $F_{EXT}(y_A)$ is the percentile of the historical average latent variable of individual loans at an extreme location and refers to the confidence demanded.

Recall that both variables y_A, one for each loan, used to calculate the probability are equal to each other because the segment/portfolio is assumed to be homogenous, so the percentile of the average Y in the extreme distribution ($= F_{EXT}(y_A)$) is the same for all loans. Consequently, the extreme percentile of PD ($=1 - F_{EXT}(y_A)$) is also the same for all loans.

$\hat{C}$ is assumed to be a Clayton Copula to detect the supposed lower-tail dependence of the latent variables: they are more related in downturns when their levels are lower. For this particular case, the Clayton Copula with parameter θ is:

$$\hat{C}(F_{EXT}(y_A), F_{EXT}(y_A)) = [F_{EXT}(y_A)^{-\theta} + F_{EXT}(y_A)^{-\theta} - 1]^{-1/\theta} = [2 * F_{EXT}(y_A)^{-\theta} - 1]^{-1/\theta}$$

This formula gives the probability of the latent variable being jointly smaller than its historical average when the latter reaches an unusually high percentile, $F_{EXT}(y_A)$, in the respective distribution. This corresponds to the likelihood of losses being simultaneously above an extreme point and the expression above substitutes $(K_V - PD)$ in [5.1]. Therefore the capital to cover unexpected losses is:

$$LGD * [(2 * F_{EXT}(y_A)^{-\theta} - 1]^{-1/\theta})] * Maturity \qquad [\ 5.9\ ]$$

where LGD and $Maturity$ are defined as in [5.1], $F_{EXT}(y_A)$ is the extreme percentile of the average latent variable calculated in [5.8] according to the confidence required, and θ is the parameter of the Clayton Copula, estimated from the rank correlation (Kendall's tau or Spearman's rho) of PD (following, respectively, [2.1] and [2.2] from Chapter 2).

5.4.3 Additional comments on this alternative model

A prior use of copulas in order to suggest some improvements to Basel II was reported in Benvegnù et al. (2006). The main purpose was to capture diversification effects, since the Basel II determines the simple addition of all capital requirements for segments without taking correlations into account. Their analysis was focused on corporate loans and concluded that the copula approach reduces the capital required by 10 to 30%.

However, *“to be in line with the model used in the Basel II credit framework and the major industry models”* (p. 497), the authors assumed that the loans have Vasicek distributions, the underlying factors that drive credit losses were jointly normally distributed, and the dependence between them was also Normal (Gaussian Copula). Such assumptions restricted the identification of simultaneous extreme occurrences.

Here, the relation between the latent variables is assumed to be satisfactorily represented by the Clayton Copula in order to find their lower-tail dependence (i.e. lower levels of latent variables, which lead to defaults, are more correlated in lower economic levels).

In short, although copulas enable us to capture the diversification effects among different segments (which tend to reduce the capital necessary to cover unexpected losses, as in Benvegnù et al., 2006) some of their families identify higher level of dependence at the extremes (which may increase the capital needed). Thus, due to the assumption of tail dependence, the proposed formula in this section is more conservative and it is aligned with regulators’ point of view (and practitioners who want to guarantee adequate capital to cover losses in severe scenarios).

It could be said that if “real” data do not present intense tail dependence, the capital calculated by [5.9] will be excessive. But if there are chances of overestimation, regulators and institutions that adopt this approach may reduce the confidence of the extreme average latent variable used as an input in the formula. Even in this case, the alternative method seems to be more appropriate than the current Basel Accord given that the latter assumes an unrealistic distribution for the variables involved and measures the dependence between them by using the linear correlation coefficient which does not capture tail dependence.

Furthermore, the copula-based approach has other advantages: it may be used for negatively correlated losses (provided that the rank correlation is positive) while Basel II’s model does not admit negative correlation; and it does not assume any specific type of distribution for credit losses, the latent variable, and the unobserved economic factor.

5.5 SIMULATIONS AND RESULTS FOR REQUIRED CAPITAL

Simulations were used to test the efficiency of the alternative model. The capital required according to the Basel method was computed for three types of consumer loans (to which the maturity adjustment is not applied)[49]: revolving credit, mortgage, and "other retail". For simplicity, LGD was assumed equal to 100% (i.e. the Recovery Rate is 0%).

The simulations were controlled for three variables, PD (15 rates between 1% and 15%, inclusive), PD dependence expressed by the Gumbel Copula's parameter[50] (11 values from 1.05 to 2), and the shape of PDs' distributions (normal/gaussian, exponential, beta, and gamma) which represented 660 scenarios. Apart from the case of normal PDs, the other three distributions were simulated in such ways that their parameters resulted in the mean (PD) chosen and in distributions skewed to the right indicating asymmetric high losses (following Kalyvas et al., 2006 who stated that credit losses present distributions skewed to the right).

Note that the selection of the Gumbel Copula implies the existence of upper-tail dependence for losses. The higher the parameter, the higher that dependence. The confidence[51] of the proposed model was set at 0.90. Each scenario contained 1,000 observations (equivalent to 1,000 periods) and was run 1,000 times to minimise possible randomness effects on results[52].

---

[49] These simulations can be run for corporate debt as well but some scenarios for the maturity adjustment should also be defined.

[50] The smallest value allowed for the Gumbel parameter is 1 (which represents independence).

[51] Other confidence levels were tested (not displayed here) but yielded lower ratios of outperformance over Basel II, mainly due to overestimations of the copula-based method.

[52] The results presented in Table 5.2 are the averages of each variable simulated. Furthermore, the codes for data generation include some commands to guarantee that the loss dependence's parameters are close enough to the stated values (divergence no greater than 0.01).

To calculate the “true” joint unexpected losses, two “correlated” variables (probabilities of default = PDs) were simulated with the same features (mean, distribution’s family, and its parameters) since the segment or portfolio in terms of calculation of capital is assumed to be homogenous. This pair of variables represents all pairs of dependent loans (all pairs have the same dependence) in the simulation criteria. Then we computed the maximum loss when the variables were simultaneously above the mean (average PD).

The performance of the models was measured according to the magnitude of the ratio between the “true” maximum unexpected losses and the capital estimated (without taking into account if the capital was excessive or insufficient to cover the losses). Thus, for instance, if the real maximum unexpected losses were 20%, one particular method resulted in 25% and other method estimated 16%, the latter was considered better because the magnitude of its divergence (= 1 - (0.16/0.20) = 0.20 deficient) was less than the difference generated by the former model (= (0.25/0.20) - 1 = 0.25 in excess).

Considering all 660 scenarios, Basel estimations for the three categories of consumer loans were concurrently better than the alternative model’s results in 26.52% of the cases. On the other hand, the copula approach was more efficient than traditional calculations for the three (at least one of the) consumer credit classes in 33.79% (73.48%) of the cases. However these ratios rise to 45.05% (92.32%) if the normally distributed losses are excluded. Therefore the performance of the copula-based method was directly related to the shape of the marginal loss distributions.

Table 5.1 – Panel A presents, for each of the loss distributions studied, the proportion of scenarios in which the alternative method gave better performance than the Basel approach did. The forecasts pertaining to exponential (normal) losses presented the best (worst) results. However, it was noticed in other simulations (not displayed here) that the results for normal PDs could be improved if lower levels of confidence were employed.

Table 5.1 – Proportion of successful estimations of the alternative method compared to Basel II estimations for consumer loans

Loss distribution	Revolving credit	Mortgage	"Other retail"	Three classes	At least one class
Panel A: All scenarios simulated					
Normal	0.00%	16.97%	3.64%	0.00%	16.97%
Exponential	100.00%	84.85%	90.91%	78.79%	100.00%
Beta	96.36%	38.79%	95.15%	38.79%	98.79%
Gamma	73.94%	17.58%	74.55%	17.58%	78.18%
Average	67.58%	39.55%	66.06%	33.79%	73.48%
Panel B: Scenarios with correlation lower than or equal to 0.16					
Normal	0.00%	43.33%	10.00%	0.00%	43.33%
Exponential	100.00%	93.33%	100.00%	93.33%	100.00%
Beta	86.67%	33.33%	73.33%	33.33%	86.67%
Gamma	66.67%	33.33%	53.33%	26.67%	66.67%
Average	50.67%	49.33%	49.33%	30.67%	68.00%

As for the classes of loans, revolving credit and "other retail" had superior performance: they were better than Basel II in around 68% and 66% of the scenarios, respectively (these figures go up to 90% and 87% if normal losses are not taken into account). The Basel formula for mortgage was more accurate because the correlation for this group is, in general, higher and this avoided excessive underestimation in some circumstances.

So, if the assumptions followed to generate the scenarios are valid for "real" portfolios, the alternative approach is liable to outperform Basel II especially for revolving credit and "other retail" whose losses are not normally distributed.

A special warning about Basel results is the high percentage of underestimated maximum potential losses: 85% with respect to revolving credit and "other retail" and 61% in mortgage portfolios. Typically, this drawback happened for non-normal losses.

As an additional analysis to get results closer to what financial institutions might experience in practice, the comparison was limited to levels of dependence of PDs that are likely to be more representative of empirical credit portfolios. The proxy for the dependence of "real" consumer loans is based on the values inferred in Basel II (from 0.03 to 0.16) for the (linear) correlation across assets of retail debtors[53]. The outperformance proportion of the alternative model with regard to portfolios correlated in that restricted range is displayed in Table 5.1 – Panel B.

On average, the alternative approach yielded worse results for portfolios less correlated (Panel B compared to Panel A). This result is explained by the fact that the main benefit of using the Clayton Copula method is the identification of left tail dependence and the consequent higher number of joint occurrences in the extreme left side of the latent variable's distribution (or, equivalently, in the right tail of the loss distribution). Since loans presenting lower correlation (as those in Panel B of Table 5.1) tend to have reduced degree of tail dependence, the poorer performance of the suggested model in these cases was expected.

Table 5.2 shows some examples[54] of capital estimated using the copula technique and Basel II for consumer portfolios with correlation compatible with the values adopted by the Basel Committee on Banking Supervision (BCBS) – see the fourth column[55]. The maximum unexpected losses observed in the simulated portfolios and the best approximations are highlighted in boldface.

If regulators and/or practitioners wish to set particular dependence values for each type of loan instead of calculating them directly from every single portfolio, the copula model can still be used successfully through the definition of a copula parameter for each credit category (which can be inferred from rank correlations between losses – that reflect the rank dependence across debtors' assets – by utilising [2.1] or [2.2] presented in Chapter 2).

[53] The correlations adopted by Basel II model are: 0.04 for revolving credit, 0.15 for mortgages and from 0.03 to 0.16 (as a decreasing function of PD) for "other retail credit".

[54] Among the 15 PDs simulated, seven were selected: 0.01, 0.03, 0.05, 0.07, 0.10, 0.12, and 0.15.

[55] The correlations specified in Basel II refer to obligors' asset returns (latent variables) and are assumed to drive the correlations across PDs. Thus, we use the latter (observable in the simulations) as a proxy for the former (unobservable).

Table 5.2 – Comparison between capital calculated by Basel Model and the alternative formula for some of the simulated credit portfolios (with linear correlation between 0.03 and 0.16, inclusive)*

PD	PD dependence (Gumbel θ)	Latent variable dependence (Clayton θ)	Linear correlation	"True" maximum unexpected losses	Alternative required capital	Capital required Basel II (revolving)	Capital required Basel II (mortgage)	Capital required Basel II ("other retail")
Panel A : Normal distribution								
0.01	1.05	0.0997	0.0770	**0.0132**	0.0494	**0.0306**	0.1003	0.0814
0.01	1.10	0.1999	0.1472	**0.0141**	0.0555	**0.0306**	0.1003	0.0814
0.03	1.05	0.1010	0.0787	**0.0396**	0.1411	**0.0687**	0.1991	0.1116
0.03	1.10	0.2005	0.1475	**0.0425**	0.1529	**0.0687**	0.1991	0.1116
0.05	1.05	0.1003	0.0783	**0.0670**	0.2293	**0.0973**	0.2635	0.1181
0.05	1.10	0.2003	0.1475	**0.0701**	0.2440	**0.0973**	0.2635	0.1181
0.07	1.05	0.1029	0.0798	**0.0927**	0.3162	**0.1207**	0.3111	0.1231
0.07	1.10	0.2011	0.1473	**0.0993**	0.3317	**0.1207**	0.3111	0.1231
0.10	1.05	0.1037	0.0805	**0.1320**	0.4436	0.1491	0.3634	**0.1343**
0.10	1.10	0.2024	0.1488	**0.1406**	0.4586	0.1491	0.3634	**0.1343**
0.12	1.05	0.1038	0.0803	**0.1589**	0.5273	0.1649	0.3895	**0.1434**
0.12	1.10	0.2013	0.1479	**0.1693**	0.5407	**0.1649**	0.3895	0.1434
0.15	1.05	0.1023	0.0795	**0.1996**	0.6512	**0.1847**	0.4191	0.1575
0.15	1.10	0.1998	0.1469	**0.2115**	0.6614	**0.1847**	0.4191	0.1575
Panel B: Exponential distribution								
0.01	1.05	0.1033	0.1012	**0.0469**	**0.0496**	0.0306	0.1003	0.0814
0.03	1.05	0.1017	0.1001	**0.1446**	**0.1412**	0.0687	0.1991	0.1116
0.05	1.05	0.1006	0.0971	**0.2366**	**0.2293**	0.0973	0.2635	0.1181
0.07	1.05	0.1012	0.1014	**0.3346**	**0.3158**	0.1206	0.3111	0.1231
0.10	1.05	0.1039	0.1010	**0.4808**	**0.4438**	0.1492	0.3634	0.1343
0.12	1.05	0.0984	0.0973	**0.5706**	**0.5264**	0.1649	0.3895	0.1434
0.15	1.05	0.1046	0.0991	**0.7012**	**0.6513**	0.1847	0.4191	0.1575

(continued on the next page)

(*) The maximum unexpected losses observed and the best estimation for each scenario are highlighted.

Table 5.2 (continued) – Comparison between capital calculated by Basel Model and the alternative formula for some of the simulated credit portfolios (with linear correlation between 0.03 and 0.16, inclusive)*

PD	PD dependence (Gumbel θ)	Latent variable dependence (Clayton θ)	Linear correlation	"True" unexpected losses	Alternative required capital	Capital required Basel II (revolving)	Capital required Basel II (mortgage)	Capital required Basel II ("other retail")
Panel C: Beta distribution								
0.01	1.05	0.1024	0.1040	**0.2136**	0.0496	0.0306	**0.1003**	0.0814
0.03	1.05	0.1043	0.1029	**0.2813**	0.1416	0.0688	**0.1991**	0.1116
0.05	1.05	0.1040	0.1003	**0.3149**	0.2298	0.0973	**0.2635**	0.1181
0.07	1.05	0.1029	0.0972	**0.3344**	**0.3163**	0.1207	0.3111	0.1231
0.10	1.05	0.1010	0.0947	**0.3574**	0.4432	0.1491	**0.3634**	0.1343
0.12	1.05	0.1040	0.0940	**0.3717**	**0.5272**	0.1649	**0.3895**	0.1434
0.15	1.05	0.1024	0.0917	**0.3817**	**0.6511**	0.1847	**0.4191**	0.1575
Panel D: Gamma distribution								
0.01	1.05	0.1015	0.1056	**0.1494**	0.0495	0.0306	**0.1003**	0.0814
0.03	1.05	0.1026	0.1034	**0.2027**	0.1413	0.0688	**0.1991**	0.1116
0.05	1.05	0.1025	0.0983	**0.2336**	**0.2296**	0.0973	0.2635	0.1181
0.07	1.05	0.1050	0.0993	**0.2654**	0.3164	0.1206	0.3111	0.1231
0.10	1.05	0.1012	0.0957	**0.2980**	0.4433	0.1492	**0.3634**	0.1343
0.12	1.05	0.1041	0.0955	**0.3126**	**0.5274**	0.1649	**0.3895**	0.1434
0.15	1.05	0.0985	0.0907	**0.3383**	**0.6508**	0.1847	**0.4191**	0.1575

(*) The maximum unexpected losses observed and the best estimation for each scenario are highlighted.

5.6 CONCLUSIONS

Due to the assumptions of normally-distributed variables and the use of a linear measure of dependence (correlation coefficient), Basel method is not able to identify joint extreme events accurately. Therefore, the capital demanded to cover unexpected losses may be misestimated.

The main contribution of this study is to consider potential tail dependence between associated (latent) variables to calculate the probability of credit losses in adverse situations. By capturing joint extreme events more precisely by means of diverse dependence structures (copulas), the alternative model improves the accuracy of estimates related to simultaneous large losses which usually happen in downturns.

Differently from the studies that estimate losses at the portfolio level based on the assumption of normality, such as Vasicek (1991 and 2002), Gordy (2003) and BCBS (2005), the method introduced in this chapter does not assume any specific distribution for the variables considered. Our approach is more flexible than the method adopted in Benvegnù et al. (2006) who were limited to the Gaussian Copula whilst we set a framework compatible with any copula family.

The formulas proposed here can be easily implemented and are intended to replace the term in Basel II referent to the subtraction of the extreme default rate (K_V) by PD (see [5.1]). Nevertheless, some basic assumptions of the Basel II approach are kept, namely: the homogeneity of segments/portfolios and the fact that defaults are driven by latent variables which are impacted by an unobserved (economic) factor. Also, possible pitfalls related to the calculation of the loss given default (LGD) and the maturity adjustment are not investigated.

Simulations of right-tail-dependent losses that controlled for several levels of PDs, their dependencies and marginal distributions confirmed the superiority of the suggested method when losses are not normally distributed. Hence, given that the literature has presented some evidence that credit losses do not follow the normal distribution and have tail dependence, the copula-based model is likely to outperform the current method in many (or most) of the loan portfolios held by financial institutions.

Even if the dependence structure adopted in the exemplary model (Clayton Copula) is considered too rigorous, it still can be used without major concerns if the confidence is reduced.

Naturally, the good performance of the alternative model shown for some scenarios in Section 5.5 is valid only if losses have upper-tail dependence. The next step to consolidate the application of this approach is the empirical search for the copula family and respective parameter(s) that best represent the relationship between latent variables (which may result in different families and parameters for distinct classes of credit, such as corporate, mortgage, revolving, and so on).

Another promising extension of this study is the use of Copula Theory to evaluate another component in the Basel formula: the loss given default (LGD).

The setup introduced in this chapter to estimate extreme (unexpected) credit losses is not the only way of using copulas in the Basel Accords' context. Another model that employs copula families related to Poisson Processes will be presented in the next chapter. The second alternative approach (Poisson) is applicable to a wider range of negative correlations while the copula-based method presented in this chapter is suitable for negative correlation only when the rank dependence (e.g. Kendall's tau or Spearman's rho) is positive[56] (see Section 5.4.3). This exclusive advantage of the Poisson model may be interesting to calculate the capital to cover unexpected losses in financial institutions given that the empirical analysis in Chapter 3 showed the existence of some negatively-related credit segments.

[56] Which comprises few cases given that, in these circumstances, the rank correlation is positive only if the negative linear correlation has small magnitude (i.e. is close to zero).

CHAPTER 6
ESTIMATION OF JOINT CREDIT LOSSES BASED ON POISSON PROCESSES AND A SUGGESTION FOR BASEL ACCORDS

"Do not go where the path may lead, go instead where there is no path and leave a trail." (Ralph Emerson)

6.1 INTRODUCTION

Credit risk factor models assume that assets are impacted by systematic and specific risks and that the correlation across asset returns results from the association between latent variables (log-returns of debtors' assets) driven by the systematic portion of the risk. The formula used to represent the latent variables in these models is an application of a property of equicorrelated jointly normal variables and therefore its results are conditional on the assumption that all variables involved in the calculation are normally distributed. However there are many studies in the literature showing that asset returns do not follow the normal distribution.

The aim of this chapter is to apply Poisson processes along with copulas to estimate joint credit losses. It uses the idea that the dependence between asset performances is driven by latent variables which are interpreted as asset "lifetimes" or "time until default". By considering systematic and idiosyncratic risks as independent "fatal shocks", we can use Poisson processes to represent the arrival time of these shocks which, in turn, is equivalent to the assets' lifetimes. It should be noted that negative shocks in this context correspond to "impulses" that lengthen asset lifetimes.

The contributions of this study are twofold. First, we propose a model to estimate conjunct credit losses based on Poisson processes and related copulas. This approach has the advantages of relaxing the assumption of normality (for losses) present in traditional credit models (such as CreditMetrics® and KMV®) and of incorporating some levels of negative correlations across losses of different debtors. Second, we use the Poisson model to derive a formula to calculate the capital to be held by financial institutions to cover unexpected credit losses. In doing so, we extend the benefits of the Poisson method (non-normality and compatibility with negative correlation) to the banking regulation context.

The remainder of this chapter is split into four sections. Some concepts necessary to the development of the proposed model are presented in Section 6.2. Next, the Poisson-based model is derived and some simulations confirm its efficiency. In Section 6.4, the model is adjusted to calculate unexpected credit losses and simulations show that it outperforms Basel formula in some scenarios that may represent real credit portfolios. Section 6.5 contains some conclusions and possible extensions.

6.2 POISSON PROCESSES AND RELATED COPULAS

Poisson processes are widely used to represent the "arrival time" of independent shocks that affect components of a system. Such shocks may be non-fatal or fatal depending on whether the components survive or fail, respectively. The "waiting time" until the next shock is assumed to be exponentially distributed and the processes are characterised by an intensity parameter λ that indicates the expected number of events (shocks) per period.

Considering the case of fatal shocks, it is clear that the time of the shock represents the lifetime of the component affected. So, if the shocks are independent we can use Poisson processes to estimate components' lifetimes.

Consider a system with two components as an example (a two-dimensional Poisson process). They are subject to "shocks" that may be fatal to one or both components. This could be the case of a small factory with two machines: one of them may fail owing to a problem in that specific machine or both may stop working if the factory has a general problem.

Let T_1 be the lifetime of component 1 and T_2 the lifetime of component 2. If we are interested in estimating the probability of both components (machines, e.g.) "surviving" beyond a particular time *t* we need to calculate $Pr[T_1 > t, T_2 > t]$.

More specifically, consider two components with lifetimes T_1 and T_2 that may "suffer" three independent shocks whose times S_1, S_2, and S_{12} are exponential random variables with positive parameters (occurrence rates) λ_1, λ_2, and λ_{12} that affect, respectively, only component 1, only component 2, and both of them.

The calculation of $Pr[T_1 > t_1, T_2 > t_2]$ leads to the Marshall-Olkin Copula defined as (see Marshall and Olkin, 1967 and Nelsen, 2006, Chapter 3):

$$\hat{C}(u,v;\alpha,\beta) = \min(u^{1-\alpha}v, uv^{1-\beta}) = \begin{cases} u^{1-\alpha}v, u^{\alpha} \geq v^{\beta} \\ uv^{1-\beta}, u^{\alpha} \leq v^{\beta} \end{cases} \quad [\ 6.1\ ]$$

where $F(.)$ indicates a cumulative distribution function, $u = \overline{F}_{T_1}(t_1) = 1 - F_{T_1}(t_1)$, $v = \overline{F}_{T_2}(t_2) = 1 - F_{T_2}(t_2)$, $\alpha = \lambda_{12}/(\lambda_1 + \lambda_{12})$, and $\beta = \lambda_{12}/(\lambda_2 + \lambda_{12})$. The notation $\hat{C}$ is used to highlight the fact that this family is a Survival Copula (defined in Chapter 2, Section 2.3.1).

This copula gives the probability of $T_1 > t_1$ at the same time that $T_2 > t_2$ and when $\alpha = \beta$ (i.e. $\lambda_1 = \lambda_2$), the dependence corresponds to the Cuadras-Augé Copula (Cuadras and Augé, 1984):

$$\hat{C}(u,v;\alpha) = [\min(uv)]^{\alpha}[uv]^{1-\alpha} = \begin{cases} uv^{1-\alpha}, u \leq v \\ u^{1-\alpha}v, u \geq v \end{cases} \quad [\ 6.2\ ]$$

where the same notation of [6.1] applies.

6.3 A GENERAL MODEL TO ESTIMATE JOINT PROBABILITIES OF DEFAULT BASED ON POISSON PROCESSES

6.3.1 The model

Consider that joint defaults result from the dependence between latent variables represented by the “time until default” (lifetimes) of assets, T. The probability of default, PD, is the likelihood that a “fatal” shock will happen in a specific time (“unit time”, u, to use Poisson’s terminology):

$$PD = \frac{u}{T} \qquad [\ 6.3\ ]$$

For example, let u = 1 year and T = 20 years. In this case, PD = 0.05 indicating the probability that the default will occur in the next year. In fact, T is a latent variable and must be inferred from PD (which is observable[57]).

Note that PD and T are expected values and thus PD may oscillate as a function of T. Whenever T increases (decreases), PD decreases (increases) for a fixed unit time. Therefore the probability of PD being greater than a specific level, pd, is equal to the probability of asset lifetimes being smaller than the lifetime, t, correspondent to that loss level.

Figure 6.1 shows this equality. The shaded areas are equal to each other so that $Pr[PD>pd] = Pr[T < t]$ and therefore high values of pd are associated with low asset lifetimes (and vice versa). Bear in mind that both distributions are merely illustrative since the lifetime distribution will be defined ahead according to Poisson model specifications and the loss distribution will be kept unknown.

Also, note that if the cumulative area until pd is p, pd is the p^{th} percentile of the PD distribution, the area below t is $1 - p$ and t is the $(1 - p)^{th}$ percentile of the asset lifetime distribution.

[57] As a frequency of defaulted assets in a portfolio (in a specific period) or deduced from ratings for individual assets (debt issuers).

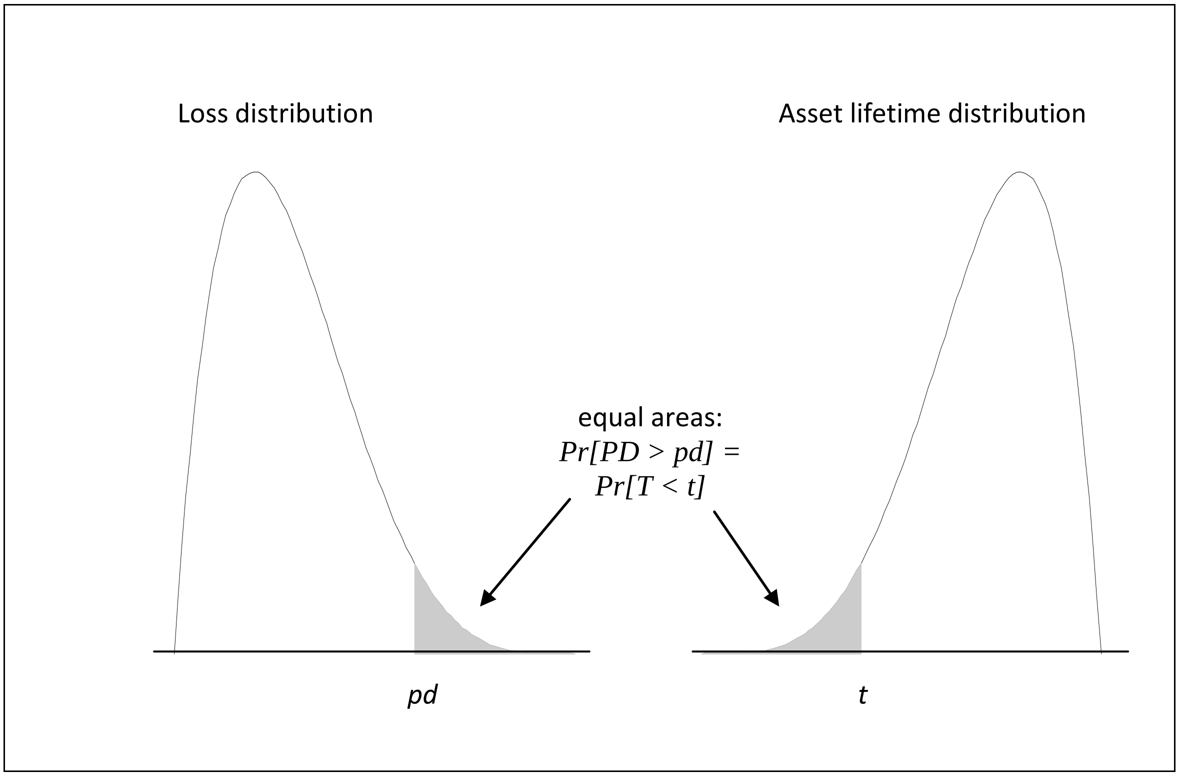

FIGURE 6.1 – Illustration on the equivalence between areas above a specific probability of default and below the associated latent variable's level (asset lifetime).

This means that the likelihood of PD being *above (below)* a specific percentile p of the loss distribution is equal to the probability of the asset lifetime being *below (above)* the percentile $1 - p$ of the latent variable distribution.

Knowing the shape of the cumulative distribution $F_{PD}(pd)$ is irrelevant; we just need to define the level of unfavourable scenarios (high PDs) we will test, which is represent by the area above pd in Figure 6.1.

Like in factor models, we assume that each asset faces systematic and independent idiosyncratic risks (shocks). Defaults will occur when these shocks are "fatal"[58]. For an asset i, T_i will be its lifetime, I_i will be the time in which the

[58] Since the focus is on defaults, non-fatal shocks (causes of downgrade of the loans, i.e. reduction of credit quality) are out of the scope of this study.

fatal idiosyncratic shock happens and S_i the time of the systematic shock. So, $T_i = min(I_i, S_i)$.

The joint probability of lifetimes of two assets i and j being shorter than a particular period t is expressed as $Pr[T_i < t, T_j < t]$. This will occur only if the idiosyncratic shocks of both assets are shorter than t or if the systematic shock happens before that time. Since the specific shocks are independent, we can infer that the probability of joint defaults caused by conjunct idiosyncratic reasons is negligible for large portfolios in short intervals of time. It is not plausible that many companies will be affected by, e.g., "bad management" at the same time (in the case of corporate debt) or people will have personal problems at once (in the case of retail loans). Therefore, the analysis of *joint* default in portfolios should concentrate on the systematic portion of the risk.

We are concerned about situations in which several assets present "high" PDs simultaneously but we are not able to calculate these joint occurrences since we do not know the loss distribution. Based on all prior assumptions, it is reasonable to assume that shock times (and therefore assets' lifetimes) follow Poisson processes implying that those variables are exponentially distributed. Thus, we can find the joint probability of "low" potential asset lifetimes and use it as a proxy for the joint probability of "high" PDs. In the case of two assets, for example, we should calculate $Pr[T_i < t, T_j < t]$, i.e. the probability that the lifetimes of assets i and j will be smaller than t.

For assets i and j, the Marshall-Olkin copula, $\hat{C}_{T_{i,j}}$, gives $Pr[T_i > t_i, T_j > t_j]$ but we are interested in $Pr[T_i < t_i, T_j < t_j]$. As the bivariate survival copula $\hat{C}(\overline{F}_{T_i}(t_i), \overline{F}_{T_j}(t_j)) = \Pr[T_i > t_i, T_j > t_j]$ can be written as $1 - F_{T_i}(t_i) - F_{T_j}(t_j) + \Pr[T_i < t_i, T_j < t_j]$, we have:

$$F_{T_{i,j}}(t_i, t_j) = \Pr[T_i < t_i, T_j < t_j] = F_{T_i} + F_{T_j} + \hat{C}_{T_{i,j}} - 1 \qquad [\,6.4\,]$$

where $F_{T_{i,j}}$ is the joint distribution of T_i and T_j evaluated at t_i and t_j.

F_T, which represents the respective marginal (exponential) distributions of lifetimes of assets i and j, is given by $1-e^{-(\lambda_{idio}+\lambda_{syst})t}$ where λ_{idio} and λ_{syst} are the intensity ("expected" occurrences) of idiosyncratic and systematic fatal shocks, respectively.

So, the complete expression, derived from [6.4] using [6.1], is:

$$F_{T_{i,j}}(t_i, t_j) = 1 - e^{-(\lambda^i_{idio} + \lambda_{syst})t_i} + 1 - e^{-(\lambda^j_{idio} + \lambda_{syst})t_j} + \quad [6.5]$$

$$\min(e^{[-(\lambda^i_{idio}+\lambda_{syst})t_i][1-(\lambda_{syst}/(\lambda^i_{idio}+\lambda_{syst}))]}.e^{-(\lambda^j_{idio}+\lambda_{syst})t_j}, e^{-(\lambda^i_{idio}+\lambda_{syst})t_i}.e^{[-(\lambda^j_{idio}+\lambda_{syst})t_j][1-(\lambda_{syst}/(\lambda^j_{idio}+\lambda_{syst}))]}) - 1$$

where $1 - e^{-(\lambda^i_{idio} + \lambda_{syst})t_i} = F_{T_i}$, $1 - e^{-(\lambda^j_{idio} + \lambda_{syst})t_j} = F_{T_j}$, and the term $min(.)$ corresponds to the copula $\hat{C}_{T_{i,j}}$ in [6.4], in this case, a Marshall-Olkin Copula associating $\overline{F}_{T_i} = 1 - F_{T_i} = e^{-(\lambda^i_{idio} + \lambda_{syst})t_i}$ to $\overline{F}_{T_j} = 1 - F_{T_j} = e^{-(\lambda^j_{idio} + \lambda_{syst})t_j}$.

t (which is exponentially distributed) is the time to be specified according to the confidence demanded (area below point t in the lifetime distribution). It may be found by $F^{-1}(confidence, 1/\lambda)$, where $F^{-1}(.)$ stands for the inverse exponential distribution and confidence is a value in [0,1].

In this model, we consider only cases where $T \geq u$, i.e. the asset lifetime is not allowed to be smaller than the period analysed. So, for example, if we are studying the probability of default in the next two years (u = 2), the shortest expected asset lifetime T is 2. Based on [6.3], this constraint implies that $\lambda \in (0,1]$ for u > 0. For instance, in the example given just after [6.3], when u = 1 year and T = 20 years, we expect a "0.05 shock" in year 1 (i.e. λ = 0.05). Therefore λ is usually non-integer (the only exception is λ = 1 when $T = u$) and corresponds to the probability of default (PD) of an asset in the period specified. This fractional λ does not affect the use of the Poisson model given that there is no requirement concerning integer intensity parameters. Note that the default for *each* asset may be caused by a specific shock (represented by λ_{idio}) and/or a general shock (λ_{syst}). As $\lambda \in (0,1]$, this means that both λ_{idio} and λ_{syst} must also be in the range (0,1].

We do not need to distinguish λ_{idio} and λ_{syst} because we are interested in the PD itself (which is equal to $\lambda = \lambda_{idio} + \lambda_{syst}$) regardless of its cause but knowing (or assuming) the ratio $\lambda_{syst} /(\lambda_{idio} + \lambda_{syst})$ is essential to calculate the joint default probability. This term indicates the proportion of conjunct defaults caused by systematic shocks and corresponds to the copula parameters presented in [6.1] and [6.2].

As said before, virtually all joint defaults in "large" portfolios in a short period are triggered by systematic factors. Thus all simultaneous credit losses will reflect a certain association among assets (the copula parameter, in this case). Since both copulas are not continuous, we cannot apply the maximum likelihood techniques described in Chapter 3, Section 3.2.1 (because those techniques use density functions and therefore demand the derivation of the copulas). Thus, the ratio between λ_{syst} and (λ_{idio} + λ_{syst}) will be approximated here by measures of dependence across assets' losses that are in the same range of the possible values of the copula parameters. In the simulations ahead, two of these measures will be tested: the linear correlation coefficient and the Kendall's tau (rank correlation). We do not intend to say that these measures are mathematically equal to the copula parameters; they are just used as approximations due to the difficult in estimating the copula parameters.

In principle, both dependence measures used as proxies for the parameters of the Marshall-Olkin and the Cuadras-Augé copulas should be restricted to nonnegative values, i.e. in [0,1]. However the Poisson-based model can be extended to "negative systematic shocks" (negative λ_{syst}) and therefore has the advantage of being compatible with some values of negative correlations in view of the fact that when the ratio λ_{syst} /(λ_{idio} + λ_{syst}) is negative the shocks act as impulses to assets' lifetimes. However, since we use dependence measures such as the linear correlation and the rank correlation as proxies for the ratio λ_{syst} /(λ_{idio} + λ_{syst}) and these dependence measures are in the range [-1,1], the model is valid for negative λ_{syst} only when $\lambda_{idio} \geq 2 * |\lambda_{syst}|$ where |.| represents an absolute value and, due to this constraint, the lowest negative value allowed for λ_{syst} is -0.50 (given that $0 < \lambda_{idio} \leq 1$). This entails the extension of the copula parameters in [6.1] and [6.2] (which are usually taken as positive in practical applications) to some possible negative values. This change is compatible with Copula Theory given that the three basic properties of copulas mentioned in Chapter 2, Section 2.3.1, are also satisfied for the range of negative parameters (represented by the correlation coefficient or the rank correlation) mentioned above (i.e. when $\lambda_{idio} \geq 2 * |\lambda_{syst}|$).

In this fashion, when $\lambda_{syst} < 0$ and the condition $\lambda_{idio} \geq 2 * |\lambda_{syst}|$ is satisfied, opposite shocks may offset one another provided that PD (= λ = λ_{idio} + λ_{syst}) is kept in the interval [0,1], i.e the total shocks (= λ_{idio} + λ_{syst}) are still positive and comply with the Poisson process condition related to the parameter being positive. This is the case when economic conditions (systematic shocks) are so favourable that they reduce the effects of individual shocks[59]. Thus the best economic scenarios in terms of risk happen when assets are negatively correlated (although this is seldom observed in reality). Note that, in this interpretation of Poisson processes at the *portfolio* level, λ_{idio} and λ_{syst} are not individually seen as Poisson parameters. They are part of the effective parameter λ which must be positive (implying that PD will be also positive). Hence one of those two parts can be negative while the other part is positive with a greater magnitude such that $\lambda_{idio} + \lambda_{syst} > 0$.

[59] Since $PD = \lambda_{syst} + \lambda_{idio}$, if $\lambda_{syst} < 0$, $PD < \lambda_{idio}$ meaning that not all individual shocks result in default in these favourable scenarios.

After these replacements and using the correlation coefficient (ρ) as an example of a proxy for the ratio between systematic and total shocks, [6.5] becomes more intelligible:

$$F_{T_{i,j}}(t_i, t_j) = 1 - e^{-(pd_i)t_i} - e^{-(pd_j)t_j} + \min(e^{[-(pd_i)t_i][1-\rho_{ij}]}.e^{-(pd_j)t_j}, e^{-(pd_i)t_i}.e^{[-(pd_j)t_j][1-\rho_{ij}]}) \quad [6.6]$$

Moreover, note that $e^{-(pd)t}$ is uniformly distributed in (0,1) since t is exponentially distributed with parameter pd. Such a uniform distribution gives the area above the selected quantile of t in the exponential distribution. Denoting this cutoff quantile as $F_T(t)$, we can find t by employing the formula of the inverse exponential distribution $t = -\frac{1}{pd}\ln(1 - F_T(t))$ and therefore we conclude that $e^{-(pd)t} = 1 - F_T(t)$

Recall that $F_T(t_i) = 1 - F_{PD}(pd_i)$ and thus $Pr[T < t_i] = Pr[PD > pd_i]$. In other words, this means that the area below t_i in the lifetime distribution is equal to the area above the associate pd_i in the PD distribution.

So, if $F_{PD}(pd_i)$ and $F_{PD}(pd_j)$ give the default cutoffs (quantiles) for assets i and j, $e^{-(pd_i)t_i} = 1 - F_T(t_i) = F_{PD}(pd_i)$ and $e^{-(pd_j)t_j} = 1 - F_T(t_j) = F_{PD}(pd_j)$ which transforms [6.6] into:

$$F_{T_{i,j}}(t_i, t_j) = \overline{\overline{F}}_{PD_{i,j}}(pd_i, pd_j) = 1 - F_{PD}(pd_i) - F_{PD}(pd_j) + \min(F_{PD}(pd_i)^{[1-\rho_{ij}]}.F_{PD}(pd_j), F_{PD}(pd_i).F_{PD}(pd_j)^{[1-\rho_{ij}]}) \quad [6.7]$$

This formula gives the likelihood that defaults of two assets, i and j, will be simultaneously above their respective quantiles, $F_{PD}(pd_i)$ and $F_{PD}(pd_j)$, which is equivalent to the joint probability of their latent variables (asset lifetimes) being below $1 - F_{PD}(pd_i) = F_T(t_i)$ and $1 - F_{PD}(pd_j) = F_T(t_j)$, respectively. For instance, it can give the likelihood that i's and j's PDs will be above their respective 90th historical percentiles at the same time ($F_{PD}(pd_i) = F_{PD}(pd_j) = 0.90$). In this case, their lifetimes will be below their respective 10% worst (smallest) historical values ($1 - F_{PD}(pd_i) = 1 - F_{PD}(pd_j) = 0.10$).

The correlation is allowed to be negative in the cases where $F_{PD}(pd_i) + F_{PD}(pd_j) - 1 \leq \min(F_{PD}(pd_i)^{[1-\rho_{ij}]}.F_{PD}(pd_j), F_{PD}(pd_i).F_{PD}(pd_j)^{[1-\rho_{ij}]})$, for $0 < F_{PD}(pd) < 1$. In practical terms, stronger negative correlations are compatible only with smaller PD quantiles. If we are, for instance, estimating the likelihood of joint losses over the same level when negative dependence reaches its highest intensity, $\rho = -1$, consistent results will be possible just for quantiles $F_{PD}(pd)$ no greater than 0.6180.

Although Poisson processes have been used to model credit risk, to the best of our knowledge, the approach suggested in this paper to estimate joint losses is novel.

The CreditRisk$^+$ model, for example, assumes that the probability of default follows the Poisson distribution but the method adopted to derive losses at the portfolio level is different from the technique proposed here and does not employ the concept of copula (see CSFBI, 1997 and Crouhy et al., 2000).

Lindskog and McNeil (2001) suggest the application of Poisson processes to model credit risk but they focus on different questions, especially the impact of the ratio of idiosyncratic and systematic shocks on the tail of the total loss distribution.

6.3.2 Simulations

In order to test the performance of the suggested approach, we simulated normal variables with three dependence structures represented by three copulas: Gaussian (symmetric without tail dependence), Student t (symmetric with fat tails), and Gumbel (asymmetric with right-tail dependence) and calculated the probability of extreme joint occurrences. The purpose is to check whether this alternative model yields results at least as good as estimations that assume normal margins and dependence (Gaussian Copula) even when the losses are normally distributed.

Two assets, i and j, were simulated with probabilities of default equal to 0.05 and 0.10.

For each dependence structure, five strength levels were tested[60]: 0.1, 0.2, 0.3, 0.4, and 0.5 for Gaussian and Student t Copulas and 1.1, 1.3, 1.5, 2.0, and 2.5 for Gumbel Copula[61] .

Then we calculated the proportion of *joint* occurrences of PDs above specific loss percentiles: 75th for both assets, 90th for i and 80th for j, and 99th for both assets. In other words, we calculated the ratio of simultaneous losses above the mentioned percentiles of each asset loss distribution. So, for example, in the second case, we estimated the probability of i's loss being higher than 90% of its historical losses at the same time that j presents losses higher than 80% of its historical level.

Next, the joint probabilities for the same cutoffs were estimated by using two methods: (i) assuming Normal loss distributions and Gaussian dependence and (ii) the model based on Poisson Processes and using the linear correlation as proxy for the intensity of systematic shocks. Each simulation was run 10,000 times. The results are shown in Table 6.1.

We also calculated the absolute difference between each method and the real joint occurrences observed (see two last columns of Table 6.1). Such differences were used to check if, when the losses are normally distributed, the suggested Poisson method yields results as satisfactory as the estimation that

[60] These levels correspond to the copulas' parameters. The Student t Copula was simulated with degree of freedom = 1 in order to present a considerable difference from the Gaussian dependence.
[61] The parameters for Gumbel Copula are different because the minimum parameter admitted for this copula is 1 (which indicates independence).

assumes normal distributions and dependence. Based on the absolute difference, the alternative approach outperformed the traditional estimation in 21 (out of 45) scenarios (46.67%).

Naturally, this performance was conditional on the dependence structure. For the Gaussian Copula (which is the assumption of traditional estimations), the Poisson-based method was the best only in one scenario (out of 15). For the Student t and Gumbel Copulas, the suggested model gave results closer to the real observed occurrences: 13 (86.67%) and 7 (46.67%) cases out of 15, respectively.

Table 6.1 – Joint losses estimated by traditional models (assuming Normal distributions and dependence) and alternative (Poisson-based) model. Probability of losses above the specified cutoffs for two assets *i* and *j* with *PDs* = 0.05 and 0.10

Dependence strength	Cutoff *i*	Cutoff *j*	Real joint losses	Estimation assuming normality	Alternative estimation	Absolute difference: Real - normal	Absolute difference: Real - alternative
Panel A : Gaussian dependence (symmetric, no tail dependence)							
0.1	0.75	0.75	0.0729	0.0740	0.0790	0.0011	0.0061
0.2	0.75	0.75	0.0837	0.0847	0.0958	0.0010	0.0122
0.3	0.75	0.75	0.0950	0.0962	0.1132	0.0012	0.0181
0.4	0.75	0.75	0.1072	0.1083	0.1311	0.0011	0.0239
0.5	0.75	0.75	0.1203	0.1213	0.1495	0.0010	0.0291
0.1	0.9	0.8	0.0252	0.0263	0.0276	0.0011	0.0024
0.2	0.9	0.8	0.0308	0.0320	0.0353	0.0012	0.0045
0.3	0.9	0.8	0.0372	0.0383	0.0432	0.0012	0.0060
0.4	0.9	0.8	0.0439	0.0453	0.0510	0.0013	0.0071
0.5	0.9	0.8	0.0514	0.0528	0.0589	0.0014	0.0075
0.1	0.99	0.99	0.0002	0.0020	0.0011	0.0018	0.0009
0.2	0.99	0.99	0.0003	0.0021	0.0021	0.0017	0.0017
0.3	0.99	0.99	0.0005	0.0023	0.0031	0.0018	0.0025
0.4	0.99	0.99	0.0008	0.0026	0.0040	0.0018	0.0032
0.5	0.99	0.99	0.0013	0.0030	0.0050	0.0017	0.0038

Table 6.1 (continued) Joint losses estimated by traditional models (assuming Normal distributions and dependence) and alternative (Poisson-based) model. Probability of losses above the specified cutoffs for two assets *i* and *j* with *PDs* = 0.05 and 0.10

Dependence strength	Cutoff *i*	Cutoff *j*	Real joint losses	Estimation assuming normality	Alternative estimation	Absolute difference: Real - normal	Absolute difference: Real - alternative
Panel B : Student t dependence (symmetric tail dependence)							
0.1	0.75	0.75	0.0927	0.0728	0.0768	0.0199	0.0158
0.2	0.75	0.75	0.1023	0.0822	0.0916	0.0201	0.0107
0.3	0.75	0.75	0.1124	0.0922	0.1068	0.0202	0.0055
0.4	0.75	0.75	0.1234	0.1028	0.1229	0.0206	0.0005
0.5	0.75	0.75	0.1349	0.1143	0.1394	0.0205	0.0046
0.1	0.9	0.8	0.0443	0.0259	0.0266	0.0183	0.0176
0.2	0.9	0.8	0.0490	0.0307	0.0334	0.0183	0.0155
0.3	0.9	0.8	0.0541	0.0360	0.0403	0.0181	0.0138
0.4	0.9	0.8	0.0592	0.0418	0.0473	0.0174	0.0119
0.5	0.9	0.8	0.0648	0.0485	0.0546	0.0163	0.0102
0.1	0.99	0.99	0.0032	0.0019	0.0010	0.0013	0.0022
0.2	0.99	0.99	0.0035	0.0021	0.0018	0.0015	0.0017
0.3	0.99	0.99	0.0039	0.0021	0.0027	0.0018	0.0012
0.4	0.99	0.99	0.0044	0.0024	0.0036	0.0020	0.0008
0.5	0.99	0.99	0.0048	0.0027	0.0045	0.0021	0.0003
Panel C: Gumbel dependence (asymmetric tail dependence)							
1.1	0.75	0.75	0.0825	0.0791	0.0869	0.0034	0.0044
1.3	0.75	0.75	0.1123	0.1033	0.1238	0.0090	0.0115
1.5	0.75	0.75	0.1333	0.1215	0.1497	0.0117	0.0164
2.0	0.75	0.75	0.1656	0.1517	0.1881	0.0139	0.0225
2.5	0.75	0.75	0.1840	0.1703	0.2082	0.0137	0.0242
1.1	0.9	0.8	0.0331	0.0290	0.0314	0.0041	0.0018
1.3	0.9	0.8	0.0516	0.0424	0.0478	0.0093	0.0038
1.5	0.9	0.8	0.0639	0.0528	0.0590	0.0111	0.0049
2.0	0.9	0.8	0.0812	0.0702	0.0752	0.0111	0.0061
2.5	0.9	0.8	0.0895	0.0807	0.0833	0.0088	0.0061
1.1	0.99	0.99	0.0013	0.0020	0.0016	0.0007	0.0003
1.3	0.99	0.99	0.0030	0.0025	0.0037	0.0005	0.0007
1.5	0.99	0.99	0.0040	0.0030	0.0050	0.0010	0.0010
2.0	0.99	0.99	0.0057	0.0045	0.0070	0.0013	0.0013
2.5	0.99	0.99	0.0067	0.0056	0.0080	0.0011	0.0013

More interesting is that the difference between the joint losses estimated by the traditional and the suggested model is not significant (pvalue = 0.4554)[62]. Hence, we can infer that, even when loss distributions are normal, the method based on Poisson processes gives results as “good” as calculations based on assumptions of normality (dependence and loss distributions). Naturally, this implies that we could use the normality assumption with the same accuracy as the suggested method and therefore there would be no significant benefit in using the latter rather than using the former.

However recall that these simulations pertain to normally-distributed losses and they represent the scenarios in which the traditional models (based on the assumption of normality) have their best performance. So, the Poisson model was not expected to outperform traditional models in these cases.

The purpose of the simulations was to show that, *even when credit losses have the (normal) distribution assumed in traditional models, the Poisson method results in an equivalent performance*.

The simulations were repeated by using Kendall’s tau as a measure of systematic risk intensity. The results (not displayed here) in terms of performance and significance were similar to those mentioned above. However the level of outperformance of the Poisson approach according to the dependence between losses was considerably distinct: 66.67% for Gaussian, 60% for Student t, and 26.67% for Gumbel.

An additional question is whether the Poisson model outperforms traditional models when credit losses are not normally-distributed. We address this question in the next section where we apply the suggested method to estimate extreme losses in the context of capital adequacy in financial institutions and check its performance for three other loss distributions (exponential, beta and gamma).

---

[62] The difference between results of traditional and alternative methods for each dependence structure (copula) is not significant either.

6.4 A MODEL APPLICATION: CAPITAL REQUIRED TO COVER UNEXPECTED CREDIT LOSSES

6.4.1 The model

The capital calculated according to Basel II (see [5.1] in Chapter 5) is the difference between an "extreme" probability of default (PD) with a specific level of confidence (99.9%), K_V, and the average PD. An alternative formula to estimate such extreme PD may be derived from the approach presented in Section 6.3.

In [6.7], we calculated the probability of joint losses above a chosen level, $\overline{F}_{PD_{i,j}}(pd_i\ , pd_j)$. For two loans i and j, $F_{PD_{i,j}}(pd_i, pd_j)$ returns the likelihood of simultaneous PDs up to pd_i and pd_j:

$$\begin{aligned} F_{PD_{i,j}}(pd_i, pd_j) &= \overline{F}_{T_{i,j}}(t_i, t_j) = \hat{C}(1-F_T(t_i), 1-F_T(t_j)) = \\ &= \min[(1-F_T(t_i))^{1-\rho_{ij}}(1-F_T(t_j)), (1-F_T(t_i))(1-F_T(t_j))^{1-\rho_{ij}}] \end{aligned} \quad [\ 6.8\ ]$$

where $\hat{C}$ is the survival copula of lifetimes t_i and t_j, $F_T(t_i)$ and $F_T(t_j)$ are their respective quantiles (equal to one minus the quantiles of the associate PDs, i.e. $1 - F_{PD}(pd_i)$ and $1 - F_{PD}(pd_j)$), and ρ_{ij} is the correlation between default probability of loans i and j.

If the capital is stipulated for segments considered homogeneous, loans in each segment are supposed to have equal PD (and, consequently, same expected lifetime) such that the dependence between t becomes a Cuadras-Augé Copula (given in [6.2]). Thus, since $pd_i = pd_j = pd$, $t_i = t_j = t$ and substituting into [6.8], we have:

$$\begin{aligned} F_{PD_{i,j}}(pd,\ pd) &= \overline{F}_{T_{i,j}}(t,\ t) = \min[(1-F_T(t))^{2-\rho}, (1-F_T(t))^{2-\rho}] = \\ &= \hat{C}(1-F_T(t), 1-F_T(t)) = (1-F_T(t))^{2-\rho} = F_{PD}(pd)^{2-\rho} \end{aligned} \quad [\ 6.9\ ]$$

where ρ is the correlation for each pair of loans in the portfolio which are considered equicorrelated[63] and 1 – $F_T(t)$, one minus the quantile of the loans' lifetimes, is equal to the quantile $F_{PD}(pd)$ of the associated probability of default (see Section 6.3).

In this alternative approach, portfolios' losses result from joint occurrences. So, the confidence of the probability of joint defaults ("*confidencePD*") is the area under the multivariate density function up to the *PD* level chosen. Such area is equivalent to the copula $\hat{C}(1-F_T(t),1-F_T(t))$ whose range is (0,1). However, as shown ahead, the maximum consistent value for individual loss percentiles $F_{PD}(pd) = 1 - F_T(t)$ may be smaller than 1.

In principle, the *PD* distribution is unknown but the value of the *portfolio*'s lifetime T may be estimated through the inverse distribution of individual lifetimes (Exponential) with confidence (area)[64] $1-\hat{C}(1-F_T(t),1-F_T(t))$ that corresponds to the area below the lifetime t:

$$T = F_T^{-1}(confidence\,T, meanT) = F_T^{-1}(1-\hat{C}(1-F_T(t),1-F_T(t)),1/\,pd_A) = -\frac{1}{pd_A}\ln(F_{PD}(pd)^{2-\rho})$$

[6.10]

where pd_A is the average probability of default (expected loss), $F_T(t)$ is the quantile of T which is equal to one minus the quantile of the *PD* distribution ($F_T(t) = 1 - F_{PD}(pd)$) . From [6.9], $\hat{C}(1-F_T(t),1-F_T(t))$ is the confidence of *joint*

[63] Kendall's tau is an alternative to replace ρ in [6.9].

[64] The *confidenceT* is in fact given by $C(F_T(t), F_T(t))$ but, based on the prior paragraph, is approximated by 1 – *confidencePD* $=1-\hat{C}(1-F_T(t),1-F_T(t))$. This may be considered valid because loans' lifetimes in homogeneous segments/portfolios are supposed to have similar behaviour.

PD (area below specific values of PD) and $F_{PD}(pd)$ is the confidence of individual PDs. Therefore, from [6.3], when unit time = 1:

$$PD = -\frac{pd_A}{\ln(F_{PD}(pd)^{2-\rho})}$$

Since PD must be in the range [0,1], the PD quantile $F_{PD}(pd)$ must be in the interval (0, $e^{-pd_A/(2-\rho)}$] where pd_A and ρ are defined as before. So, due to this restriction in terms of individual losses' quantiles, the loss confidence[65] in this Poisson-based model will be defined as a proportion of the maximum *individual* PD quantile accepted for each specific case (i.e. a proportion of $e^{-pd_A/(2-\rho)}$).

The total *joint* losses in an extreme scenario according to a particular confidence level (= *confidence* * $e^{-pd_A/(2-\rho)}$ = proportion of the maximum individual quantile for the PD distribution) will be estimated from the value of an extreme *joint* lifetime which can be found by applying [6.10], the inverse distribution (Exponential) of *t*:

$$t_{extreme} = F_T^{-1}(confidence * e^{-pd_A/(2-\rho)}, meanT) = -\frac{1}{pd_A}\ln[(confidence * e^{-pd_A/(2-\rho)})^{2-\rho}]$$

[6.11]

Then, by using [6.3], we can calculate the extreme joint PD as:

$$pd_{extreme} = \frac{u}{t_{extreme}} \qquad [\ 6.12\ ]$$

[65] This model confidence is based on individual losses and their dependence in order to give the likelihood that each of these losses will reach a specific threshold and impact the portfolio's loss. It should not be confused with $confidenceT$ and $confidencePD$, mentioned before, which refer to the probability of *portfolios*' lifetimes and PDs, respectively, being below a particular point. If one of these two measures was used as the model confidence, the alternative method would not capture the dependence between PDs.

where the unit time u will be set equal to 1 in order to maintain the Basel II's time horizon. So, by combining [6.11] and [6.12]:

$$pd_{extreme} = -\frac{pd_A}{\ln[confidence * e^{-pd_A/(2-\rho)})^{2-\rho}]} = -\frac{pd_A}{\ln(confidence^{2-\rho}) - pd_A} \qquad [\,6.13\,]$$

with the linear correlation $\rho \in$ [-1,1], *confidence* $\in$ (0,1], and pd_A representing the average PD. As expected, $pd_{extreme}$ is increasing in *confidence* and ρ.

[6.13] gives the total probability of default in adverse scenarios with respect to the confidence required and therefore should replace the term K_V in [5.1]:

$$[LGD * (pd_{extreme} - pd_A)] * Maturity$$

Although there are different reasons for consumer defaults (Dubois and Anderson, 2010, p. 3), it is possible that this Poisson approach yields better results for consumer loans since it uses the idea of "shocks" and there is evidence that some households are more subject to shocks (such as loss of job and divorce) whilst the degradation of corporate debts is typically continuous (see Avery et. al., 2004 and Sabato, 2006).

6.4.2 Simulations

Simulations[66] of credit PD distributions were used to compare the formula based on Poisson processes to the formula determined in Basel II for three classes of retail credit (revolving, mortgage, and "other retail")[67]. 320 scenarios were created. For simplicity, LGD was assumed equal to 100%.

[66] The simulations were repeated 1000 times.
[67] Corporate credit could also be analysed but the maturity term in [5.1] should be simulated as well.

Four dependence structures (copulas) were applied to capture distinct types of association: Gaussian (symmetric without tail dependence), Student t (symmetric tail dependence), Clayton (left-tail dependence), and Gumbel (right-tail dependence).

Five levels of dependence were considered for each copula (represented by their parameters θ: 0.1, 0.2, 0.3, 0.4, and 0.5 for Gaussian, Student t, and Clayton copulas and 1.1, 1.3, 1.5, 2.0, and 2.5 for Gumbel Copula – see second footnote in Section 6.3.2). Then, we simulated four PD distributions (normal, exponential, beta, and gamma[68]) for each copula level and used four PDs in each distribution (0.01, 0.05, 0.10, and 0.15). Figure 6.2 illustrates the definition of the scenarios.

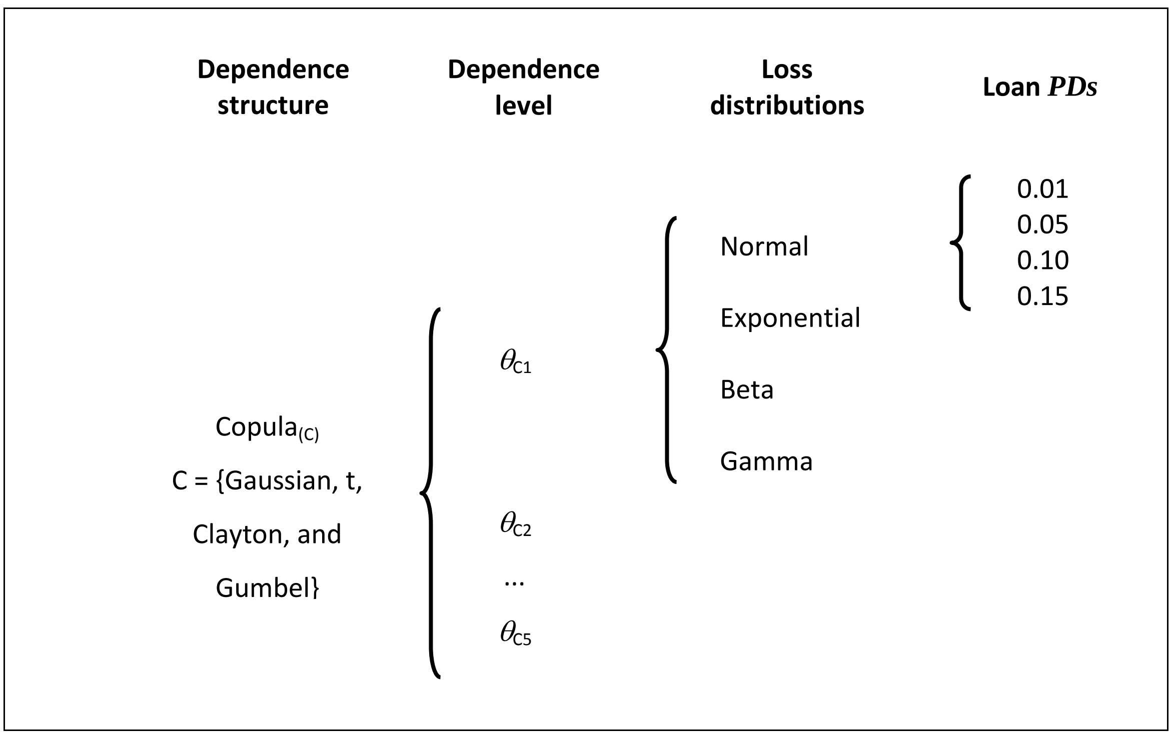

FIGURE 6.2 – Scenarios used in the simulation to compare capital to cover unexpected losses calculated according to Basel method and Poisson-based model.

[68] The parameters of the distributions were defined in such way that the mean loss was equal to the probabilities of default considered (0.01, 0.05, 0.10, and 0.15 for each distribution) and, apart from the normal case, the distributions presented “long” right tail (indicating the existence of extreme high losses).

The capital calculated according to each approach (Poisson method and Basel) was compared to the maximum unexpected losses (maximum total losses minus average PD) in all scenarios. We chose the confidence level of 95% for the alternative model[69].

The simulations revealed the superiority of the alternative formula for some scenarios. Table 6.2 shows the percentage of cases in which the alternative model gave estimations closer to the maximum unexpected losses observed in the simulated data (based on the absolute difference, i.e. without considering under or overestimations).

The performance of the suggested formula was clearly sensitive to the PD distribution. Based on the parameters used in the simulations, the Poisson model basically did not yield better results than Basel for normally distributed losses (mortgage loans represented exceptions in few cases). For the other three distributions, on average, the alternative method outperformed Basel in more than half of the scenarios (reaching 100% in some cases). The best estimations were for Gamma distributions followed by Beta and Exponential.

Changes in the confidence level improved the performance of the alternative method for some loss distributions but worsened its results in other cases. For example, if a confidence level around 0.80 is used, the Poisson method outperforms Basel II for normally distributed losses but does not yield good estimates for the other three distributions (this confidence level was tested but the results are not displayed here). However it is important to note that these results are valid only for the specific parameters used to simulate the losses.

The dependence structure does not seem to have any impact in the success of Poisson estimations since, on average, the percentage of outperformance of this suggested model was quite similar for all copulas tested.

---

[69] Among other random confidence levels tested, this value resulted in the best performance of the Poisson-based model. Recall that Basel uses the confidence of 99.9%.

From the last column of Table 6.2, we calculated that the alternative approach was simultaneously better than all three classes of retail credit in 43% of the scenarios simulated.

Table 6.2 - Percentage of Poisson-based method's estimations better than Basel II results (retail credit) using correlation coefficient calculated from the simulated data*

Loss Distribution	Revolving Credit	Mortgage	Other retail	Three classes
Panel A: Gaussian dependence				
Normal	0%	20%	0%	0%
Exponential	75%	70%	75%	50%
Beta	100%	20%	95%	20%
Gamma	100%	75%	95%	75%
Panel B: Student t dependence				
Normal	0%	20%	0%	0%
Exponential	75%	55%	75%	50%
Beta	100%	75%	100%	75%
Gamma	100%	95%	100%	95%
Panel C: Clayton dependence				
Normal	0%	25%	0%	0%
Exponential	75%	75%	55%	50%
Beta	100%	0%	75%	0%
Gamma	100%	75%	75%	75%
Panel D: Gumbel dependence				
Normal	0%	15%	0%	0%
Exponential	75%	60%	75%	50%
Beta	100%	60%	100%	60%
Gamma	100%	90%	100%	90%

(*) All the percentages in this table are divisible by 5 because there are 20 scenarios (four PD levels times five copula parameters) for each combination of loss distributions and copula families. Thus, any possible number of occurrences (from 0 to 20) in each combination will result in a percent value divisible by 5.

Table 6.3 demonstrates some examples of the difference in the capital estimated according to Basel II formula and the alternative approach.

Table 6.3 – Selected examples of capital estimated according to Basel II and suggested model

Loss distribution	Observed maximum losses	Poisson-based capital	Revolving credit (Basel II)	Mortgage (Basel II)	Other retail (Basel II)
Panel A: Gaussian dependence					
Normal	0.0656	0.3143	0.0973	0.2635	0.1181
Exponential	0.2338	0.3089	0.0973	0.2635	0.1181
Beta	0.3081	0.3071	0.0973	0.2635	0.1181
Gamma	0.3812	0.3059	0.0973	0.2635	0.1181
Panel B: Student t dependence					
Normal	0.0764	0.3099	0.0973	0.2635	0.1181
Exponential	0.2941	0.3469	0.0973	0.2635	0.1181
Beta	0.3891	0.3257	0.0973	0.2635	0.1181
Gamma	0.4984	0.3292	0.0973	0.2635	0.1181
Panel C: Clayton dependence					
Normal	0.0583	0.3027	0.0973	0.2635	0.1181
Exponential	0.2014	0.2904	0.0973	0.2635	0.1181
Beta	0.2646	0.2880	0.0973	0.2635	0.1181
Gamma	0.3297	0.2865	0.0973	0.2635	0.1181
Panel D: Gumbel dependence					
Normal	0.0770	0.3229	0.0973	0.2635	0.1181
Exponential	0.2967	0.3331	0.0973	0.2635	0.1181
Beta	0.3875	0.3329	0.0973	0.2635	0.1181
Gamma	0.4893	0.3342	0.0973	0.2635	0.1181

Among the simulations that generated the results in Table 6.2, a PD (0.5) and the intermediary dependence level (0.3 for Gaussian, Student t, and Clayton, and 1.3 for Gumbel) were selected. Considering normally distributed defaults, the Poisson model presented the highest overestimations. For the other distributions, in most of the cases, Basel approach underestimated the maximum losses while the Poisson formula returned more accurate estimations.

It could be thought that the superiority of the suggested method in some scenarios is due to the fact that it uses correlations calculated directly from the losses (PDs) while Basel employs correlations defined *a priori* (0.04 for revolving credit, 0.15 for mortgage and between 0.03 and 0.16 as a function of PD for other retail[70]). In order to test this hypothesis, we checked the performance of the Poisson-based formula using the values for ρ stipulated in Basel II Accord (with confidence level of 95%, as before).

The results are in Table 6.4 and the alternative model's performance was similar to the case where correlations calculated from the data were used (the average of the absolute difference among values in Tables 6.2 and 6.4 is 5.5%).

As in the prior simulation (Table 6.2), the worst and the best results from the suggested model were related to normally (the smallest percentage of better performance of the Poisson model) and gamma-distributed losses (the highest percentage of better performance of the Poisson model), respectively.

The formula used in Basel Accords was also tested by applying the correlation coefficient calculated from the PD series instead of using the values stipulated in the Accord. This strategy did not yield good results (not displayed here) when compared to the Poisson model as the latter outperformed the former in 67.5% of the cases.

6.5 CONCLUSIONS

The assumption of normality and the constraint of non-negative correlations are limitations of factor models. These models are derived from the structure proposed, for example, in Vasicek (2002) where the latent variables (obligors' asset returns) that drive defaults are impacted by two types of risk (systematic and idiosyncratic). The relationship among all these variables is conveniently set by means of a formula that describes equicorrelated normal distributions – see formula [2.4] in Chapter 2.

[70] See formula in BCBS (2005, 2006).

Table 6.4 - Percentage of Poisson-based method's estimations better than Basel II's results (retail credit) using correlation coefficient determined in Basel II*

Loss Distribution	Revolving Credit	Mortgage	Other retail	Three classes
Panel A: Gaussian dependence				
Normal	0%	25%	0%	0%
Exponential	75%	75%	75%	50%
Beta	100%	20%	100%	20%
Gamma	100%	75%	100%	75%
Panel B: Student t dependence				
Normal	0%	25%	0%	0%
Exponential	100%	100%	75%	75%
Beta	100%	75%	100%	75%
Gamma	100%	75%	100%	75%
Panel C: Clayton dependence				
Normal	0%	25%	0%	0%
Exponential	75%	75%	70%	50%
Beta	100%	0%	100%	0%
Gamma	100%	75%	100%	75%
Panel D: Gumbel dependence				
Normal	0%	25%	0%	0%
Exponential	95%	95%	75%	70%
Beta	100%	65%	100%	65%
Gamma	100%	75%	100%	75%

(*) All the percentages in this table are divisible by 5 because there are 20 scenarios (four PD levels times five copula parameters) for each combination of loss distributions and copula families. Thus, any possible number of occurrences (from 0 to 20) in each combination will result in a percent value divisible by 5.

On the other hand, the latent variables in the method suggested in this chapter are assumed to be loans' lifetimes. This interpretation implies that those variables can be adequately modelled by Poisson processes and are therefore exponentially distributed. Although other papers have already used Poisson processes to study credit losses (CSFBI, 1997, and Lindskog and McNeil, 2001), the model presented here introduces the use of copulas (related to lifetimes) to measure the dependence across the latent variables. Given these properties, our model allows for some values of negative dependence and tends to result in better estimations of joint losses that are not normally distributed. Even for assets with normal credit losses, the alternative Poisson model was shown to be as good as traditional estimations that assume Normal dependence and distributions.

With regard to the suggestion for capital estimation, the alternative model estimates extreme losses by using the expected loss (average PD) and the dependence among defaults (the linear correlation, for example) as inputs. Considering the parameters and the confidence level used in the simulations (Section 6.4.2), the Poisson approach did not result in good estimations for normally distributed defaults. On the other hand, this method outperformed Basel formula in credit portfolios represented by other three distributions at the 95% confidence level (whilst Basel is calculated with confidence of 99.9%). Different results in terms of loss distributions may be reached if the confidence level is changed.

Although the proposed model has the limitation of assuming independent "shocks" *for each obligor*, it may be of interest to regulators and practitioners since its implementation is relatively easy and, according to the simulations, it is typically more efficient than the current formula adopted which tends to underestimate maximum potential losses when they are not normally distributed.

One possible extension of this research is the empirical analysis of the Poisson-based model using banks' datasets to check its adequacy to model credit risk at the portfolio level. Also, other types of dependence measures should be tested to represent the ratio of systematic shocks out of the total shocks inasmuch as the use of the linear correlation coefficient is problematic because it does not capture tail dependence.

As the quality of Poisson-method estimations for specific loss distributions is sensitive to the confidence level chosen, it should be searched a way to determine the "best" confidence level for a PD dataset even if we do not know its distribution. Such a level would *indirectly* take the default distribution into account so that the extreme *PDs* estimated will be very close to the real ones.

Moreover, further research should be conducted to investigate whether the suggested Poisson model is more suitable for modelling a specific type of credit (retail or corporate).

CHAPTER 7
CONCLUSIONS AND EXTENSIONS

*It is remarkable how much we learn about how **not** to do things when we are doing research.*

7.1 SUMMARY

The assumption of normally-distributed losses[71] and the use of linear correlation in leading credit risk models may result in inaccurate estimations of joint extreme losses. Copula Theory is an option to overcome this drawback inasmuch as copula functions use individual distributions, regardless of their shape, to return the respective joint distribution.

Copula Theory has been applied to credit risk since 2000 but there are still many gaps in the literature. One of these gaps is the lack of application of copulas to study consumer loans which represent a high proportion of banks' portfolios nowadays. Chapter 3 employs copulas to estimate the dependence across credit card loans of a UK bank. The portfolio was split into five segments according to the credit quality of borrowers and the dependence (best-fit copula) was evaluated for each pair of segments following three goodness-of-fit (GoF) tests (each of them based on two different approaches: the whole default rate distribution and the right-tail of that distribution[72]). Then, the ratio of joint high default rates observed in the Bank's portfolio at specific levels was compared to the probability of concurrent high losses estimated following the assumptions of traditional credit models (normality and the use of linear correlation) and according to the best-fit copulas.

In Chapter 4, vine copulas (combinations of bivariate copulas to give higher-dimension copulas) were employed to estimate the dependence in the credit

[71] The assumption of normality for the latent variables that drive defaults implies that the losses/returns are also presumed to be normally distributed.

[72] The goodness-of-fit tests considering exclusively the right tail of the default rate distributions (above the 75th percentile) were adopted because the main objective of this study is to estimate joint *high* losses. Hence, in these cases, the copulas found were the best-fit only to the right tail of the distributions.

card portfolio for trios of segments. The copula estimations used the GoF test found to be the most robust in the literature ("Empirical Copula") and, as in Chapter 3, based on two approaches concerning the default rate distributions (complete distributions and their right tail). Next, the occurrences of simultaneous losses in trios of segments observed in the portfolio analysed were compared to estimations derived from the multivariate normality and from vine copulas.

After checking whether the dependence in the credit card portfolio studied is divergent from the assumption of traditional models (i.e. normality), we propose, in Chapters 5 and 6, two methods to determine the capital necessary to cover unexpected credit losses in financial institutions. The current approach adopted by regulators (Basel II) comes from conventional (factor) credit models and therefore has the same limitations concerning the assumption of normality and the linear correlation. So, the use of copulas in this context enables the estimation of joint extreme losses without assuming any particular distribution and may capture potential upper-tail dependence among default rates (i.e. stronger association of high probabilities of default) that result in losses higher than estimates based on presumptions of normality.

7.2 CONCLUSIONS

Our empirical analysis based on a credit card portfolio of a large UK bank revealed that default rates present tail dependence which implies that distinct levels of losses have different levels of association. In most of the cases, the tail dependence was stronger in the right side of the default distributions indicating that, particularly in recessions, the portfolio considered is exposed to losses more severe than those estimated from traditional credit models that assume normally-distributed default rates. This result confirms quite a few studies related to other assets' dependence (corporate debts included)[73]. Another

[73] See, for example, Ang and Bekaert (2002), Di Clemente and Romano (2004), Das and Geng (2006), Patton (2006) and Ning (2010).

aspect that shows the potential inaccuracy of traditional credit models is the fact that their implicit copula (Gaussian) was the best-fit dependence for only one of the ten pairs of credit card segments analysed[74]. Even in some examples where the default rates for each of two segments were normally distributed[75] the joint distribution of the pair was not normal (i.e. the dependence was not Gaussian). From this, we infer that non-normal copula families express the dependence across credit card losses more accurately than the Gaussian does. Thus, evaluations of simultaneous high credit losses based on copulas yield better results in comparison with assessments that presume multivariate normality. The performance of the former approach is improved even more when the copulas are estimated from the right tail of the default distributions (rather than from the whole distributions, as it is usually done). However this superiority is achieved at the expense of higher levels of underestimation.

Vine copulas contribute to the prediction of conjunct high default rates in portfolios with heterogeneous dependence. The empirical investigation conducted for trios of credit card segments corroborated this conclusion given that the estimates founded on vines were closer to the observed joint extreme defaults than the approximations related to bivariate copulas and to the trivariate normal distribution[76]. In some situations, vine copulas captured right-tail dependence (high default rates more linked which represents higher potential losses at the portfolio level) not identified by pairwise copulas. This is evidence that the use of vine copulas improve evaluations of possible joint high losses and their use can help financial institutions to avoid the undervaluation of potential losses and the allocation of resources to segments that are more prone to default together (mainly in unfavourable economic circumstances).

---

[74] According to goodness-of-fit tests based on the complete default distributions. None of the pairs was represented by the Gaussian Copula when the GoF tests used only the right tail of the default distributions.

[75] Although the adequacy of the normal distribution to represent the individual distributions of default rates was based on nonstatistically-significant results of the Jarque-Bera test.

[76] When compared to the pairwise-copula analysis, the vine method also had the advantage of presenting lower levels of underestimation.

The calculation of the capital to be set aside in financial institutions to cover unexpected credit losses replicates the same assumptions of factor models (normality and linear dependence) and, consequently, is subject to the underestimation of extreme defaults. The methods suggested in this thesis are able to capture stronger dependence typical in adverse scenarios and therefore can improve the quality of these estimations when losses are not normally distributed (which, according to the literature[77], represents the reality in banks).

7.3 CONTRIBUTIONS AND IMPLICATIONS

The empirical contributions of this study are related to the novel application of Copula Theory to analyses of dependence in consumer loan portfolios and the theoretical contributions concern the proposal of alternative models to estimate the capital necessary to cover extreme credit losses in financial institutions.

The empirical innovations are fourfold. First, we apply copulas to *retail* credit whilst the literature in credit risk, such as Melchiori (2003), Bo-Chih (2004), Cherubini et. al (2004), Di Clement and Romano (2004), Hull and White (2004 and 2006), Hamerle and Rösch (2005), and Das and Geng (2006), has applied this approach only to corporate debt and derivatives. Second, apart from considering five copula families (Gaussian, Student t, Clayton, Frank, and Gumbel) that are typically checked in credit studies, such as Frey et. al (2001), Frey and McNeil (2001, 2003), Schönbucher and Schubert (2001), Bluhm et al.(2002), Daul et al. (2003), Schmidt (2003), Kang and Shahabuddin (2005), and Kostadinov (2005), we include another five families (Farlie-Gumbel-Morgenstern, Galambos, Hüsler-Reiss, Joe, and Plackett) in our tests. Third, we estimate best-fit copulas according to goodness-of-fit tests based on the right tail of the defaults' distributions while this is traditionally done with focus on the complete distributions (see, for example, Cherubini et al., 2004, Di Clement and Romano, 2004, and Das and Geng, 2006). Fourth, we use vine copulas (to express higher-dimension dependence) in credit risk analyses whilst the use of

[77] For instance, Kalyvas et al. (2006) and Rosenberg and Schuermann (2006).

this technique in finance has been basically limited to stock markets, as in Aas and Berg (2009), Aas et. al. (2009), Heinen and Valdesogo (2009) and Maugis and Guegan (2010).

All the four innovations mentioned above contributed to improvements in the estimation of joint high losses. First, in general, the comparison between the probability of simultaneous high losses calculated from the multivariate normal distribution and bivariate copulas revealed that the result from the latter technique was closer to the joint occurrences observed in the credit card portfolio studied. Second, three out of the five less-tested copulas were found to be the best representation of the dependence between credit segments in some pairs. Third, estimations of joint losses according to copulas inferred from the right tail of the defaults' distributions were closer to the observed losses than the estimations based on copulas deduced from the whole distributions. Fourth, when compared with the pairwise-copula approach, vine copulas yielded better evaluations of conjunct high losses[78] and lower level of underestimation.

As for the theoretical contributions, we suggest alternative setups to evaluate simultaneous credit losses by considering some assumptions different from those used in the traditional literature, such as Vasicek (2002) and Hull and White (2004). These traditional methods are founded on the use of a relationship among equally-correlated normal variables to represent the variables studied. We look at this problem under a different perspective and assume that the dependence across the underlying variables (regardless of their distributions) can be characterised by any copula family and is closely associated to the dependence across losses (irrespective of their distributions).

The copula model presented in Chapter 5 allows the identification of potential tail dependence among credit losses (especially the upper-tail dependence found out in the literature) while, due to the assumption of normality, the current

[78] Since the bivariate-copula method outperformed the evaluations conditional on the assumption of multivariate normality, this implies that the results from vine copulas were also better than those related to the multivariate normal distribution.

method is not able to capture stronger association when losses reach extreme levels (i.e. in downturns). Moreover the alternative model is flexible and can be used with any copula family empirically found or assumed.

Apart from relaxing the assumption of normality, the Poisson model introduced in Chapter 6 is, in some cases, compatible with negatively-correlated losses whereas the current formula adopted by regulators around the world does not admit negative correlation. Although such negative relationships are not commonly found in the financial market, this improvement seems to have practical applications given that some segments of the dataset analysed in Chapters 3 and 4 are negatively associated.

Given the aforementioned benefits, the application of copulas to estimate joint extreme losses in credit portfolios can be of interest to several sectors. Academics, for example, can revise many credit models and theories founded on the linear correlation and on the assumption of normally-distributed variables. Other studies can adopt goodness-of-fit tests based only on the tails of the distributions or consider copulas rarely employed in credit studies (including those five families contemplated in Chapters 3 and 4) in other empirical investigations. Furthermore, the original use of vine copulas in this work can motivate other analyses in the credit risk field, particularly to deal with heterogeneous dependence.

The superiority of copulas over prevalent credit methods with respect to the estimation of joint extreme losses points out that financial institutions are better off employing copulas instead of relying on approaches that presume multivariate normality and linear correlation. The adoption of copulas tends to lead to more trustworthy estimates of extreme losses and this will help banks to be prepared for potential losses, notably in adverse scenarios, and to refrain from lending excessive resources to obligors and/or segments that result in higher indices of default when the economy crashes.

The main implication of this study for policy makers is the possible improvement of the calculation of the regulatory capital to be held by financial institutions to cover unexpected credit losses. The formulas derived in Chapters 5 and 6 are more efficient than Basel model at identifying tail dependence across losses. Hence the alternative approaches can detect stronger association between losses in unfavourable conditions and help to avoid the underestimation of capital needed to offset unusual shortfalls.

Although both alternative models can be easily implemented in financial institutions, it is important to note that the main contribution of these suggestions is to *open new directions* in the search for methods that guarantee more efficient estimations of joint extreme (unexpected) credit losses. Even if regulators and banks do not adopt the formulas in the exact way they were suggested here, the setups introduced in Chapters 5 and 6 can inspire the development of models that have practical applications and result in better evaluations when compared to traditional models.

7.4 LIMITATIONS AND FURTHER RESEARCH

In this thesis, the estimations of joint high losses relate to the highest *quantiles* of the losses. That is, we are concerned about situations where the highest default rates in a specific credit segment happen at the same time as the highest default rates in other groups. It is not meant to calculate the likelihood of joint losses above specific values (e.g. simultaneous default rates greater than 3% or 5%). This is left as a future exercise regarding consumer loans and will require the estimation of the marginal distribution of each segment's losses (as done by Das and Geng, 2006 for corporate debt).

The dataset used in the empirical analyses was relatively short and therefore did not have abundant information on joint extreme losses so we could not compare estimates from multivariate normality with estimates from copulas in the very tail of the distributions (where the copula models are expected to be more advantageous). Although the size of the dataset employed in Chapters 3 and 4 represents the range of data typically available in banks, one possible extension of this study is to run tests for longer periods when more observations can be obtained. Moreover other retail credit classes (e.g. mortgage and fixed term) should be considered and the results can be compared with the findings presented here for credit cards.

With respect to the study of heterogeneous portfolios, the riskiest segment in each triplet was chosen to represent the conditioning variable in the vine construction. This was done to express the most conservative scenario (the highest likelihood of joint extreme default among the combinations for each trio) and there is no guarantee that this composition gives the best approximation to the dependence across the segments. Supplementary investigations, following for example Maugis and Guegan (2010), may identify the best arrangement of variables in the cascade structure of default rates and improve even more the estimates of simultaneous losses based on vines.

The analysis in Chapter 4 was limited to D-vines (in which none of the nodes – variables – is connected to more than two other nodes). In this case, the objective is to characterise the association between segments (taken in pairs) but it is possible to apply other vines to different problems in credit risk. Canonical vines (in which one of the nodes is linked to all other nodes), for instance, are an option to study many variables conditioned to one particular variable (e.g. some proxy for the economy). Also distinct formats of vines (apart from D-vines and Canonical vines) can be employed to express more complex dependence among credit losses. Due to the absence of work that deals with this technique in credit risk, the multiple opportunities related to vines indicate a promising topic to be explored in this field (including empirical tests for other consumer classes, corporate debt and credit derivatives).

The empirical analyses were restricted to ten (one-parameter) copulas and more families can be tested in future work so that more accurate dependence structures can be found. If there is interest in the possibly differing behaviour of the credit returns (or losses) in both tails (i.e. whether "low" *and* "high" credit returns are more associated than assumed by the multivariate normal distribution but in unlike intensity), copulas with two parameters should be taken into account since they can capture distinct tail dependencies.

The copulas used here do not capture time dependence which, according to the literature, is present in asset returns; see, for instance, Conrad and Kaul (1988), Lo and MacKinlay (1988), and Bekaert (1995). This was pointed out by Fermanian and Scaillet (2005) as a drawback of empirical studies in financial markets. According to Fermanian and Wegkamp (2004), "*the research on relevant specifications for copulas and on their time dependence is still in its infancy*" (p.1).

Nonetheless the concept of conditional copulas can help to overcome this weakness; see, for example, Patton (2002, 2006), Fermanian and Wegkamp (2004), Mendes (2005), and Palaro and Hotta (2006). This approach assumes that the parameter of the copula oscillates over time, generally following an autoregressive model. Thus the strength of the dependence varies and prior realisations affect the subsequent occurrences.

Note that, in these cases, although the parameter value fluctuates, the copula family is kept constant. Another possibility should be to consider that the dependence structure may also vary over time, i.e., the copulas estimated in different periods according to goodness-of-fit tests may be different. A complete investigation in this sense would repeat all the steps in Chapter 3 and identify a best-fit copula for each time window selected. If diverse families are found to represent the same variable (e.g. credit losses) over time, this would show that the time-varying behaviour of dependence in financial markets is more volatile than supposed by academics and practitioners.

Regarding the theoretical models suggested to improve Basel approach, their successful performance compared to the estimates from the official formula was based on simulations. To check the efficiency of these alternative methods, they should be used to estimate unexpected losses in real credit portfolios and then, after some time, the results should be compared with the losses observed in those portfolios.

Furthermore those models refer to only one term of the expression that determines the capital necessary to face unexpected losses. In order to get values more representative of all aspects involving the potential higher dependence among losses in adverse scenarios, Copula Theory should be also tested to estimate the loss given default, which is another important variable that impacts the calculation.

Pertaining to the Poisson method suggested for Basel, the utilisation of the linear correlation is a shortcoming inasmuch as it does not detect tail association. So, it is interesting to search for other options of dependence measures to denote the proportion of systematic shocks.

There is empirical evidence that the suitability of the Poisson model for each loss distribution is conditional to the confidence demanded. Hence, the next step to improve this formula is to figure out the relationship (or an approximation for it) between the confidence and the loss distribution so that we can set the confidence level as a function of a variable that represents the unknown loss distribution.

REFERENCES

Aas, K., D. Berg (2009). Models for construction of multivariate dependence: A comparison study. *The European Journal of Finance*, Vol. 15, Issues 7 & 8, pp. 639-659.

Aas, K., C. Czado, A. Frigessi, H. Bakken (2009). Pair-copula constructions of multiple dependence. *Insurance Mathematics & Economics*, 44, pp. 182-198.

Andrade, D., P. Barbetta, P. Freitas Filho, N. Zunino, C. Jacinto (2006). Using copulas in risk analysis. *Proceedings of the 2006 Winter Simulation Conference*.

Andrade, F., L. Thomas (2007). Structural models in consumer credit. *European Journal of Operational Research*, 183, pp. 1569-1581.

Ang, A., Bekaert, G. (2002). International Asset Allocation With Regime Shifts. *Review of Financial Studies*, 15, pp. 1137-1187.

Avery, R., P. Calem, G. Canner (2004). Consumer credit scoring: do situational circumstances matter? *BIS Working Paper* n. 146.

Basel Committee on Banking Supervision - BCBS (2005). An Explanatory Note on the Basel II IRB Risk Weight Functions. *Bank for International Settlements Report*.

Basel Committee on Banking Supervision - BCBS (2006). International Convergence of Capital Measurement and Capital Standards: A Revised Framework. *Bank for International Settlements Report*.

Bedford, T., R. Cooke (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32, pp. 245-268.

Bedford, T., R. Cooke (2002). Vines - a new graphical model for dependent random variables. *The Annals of Statistics*, Vol. 30, N°. 4, pp. 1031-1068.

Bekaert, Geert (1995). The Time Variation of Expected Returns and Volatility in Foreign-Exchange Markets. *Journal of Business and Economic Statistics*, Vol. 13, n° 4, pp. 397-408.

Benvegnù, S., S. Fritz, W. Paus (2006). Aligning Regulatory Capital with Economic Capital: An Alternative to Risk Weights According to Basel II. In: Ong, Michael (ed.) "*Risk Management: A Modern Perspective (Volume 1)*". Elsevier Academic, pp. 481-506.

Berg, Daniel (2009). Copula goodness-of-fit testing: an overview and power comparison. *The European Journal of Finance*, Vol. 15, Issues 7 & 8, pp. 675-701.

Bernstein, Peter L. (1996). *Against the Gods. The Remarkable Story of Risk*. New York: John Wiley & Sons.

Bluhm, C., L. Overbeck, C. Wagner (2002). *An Introduction to Credit Risk Modeling.* London: Chapman & Hall/CRC.

Bo-Chih, Lai (2004). The Application of Copula in Credit Risk Management. *Working Paper*.

Bouyé, E., V. Durrleman, A. Nikeghbali, G. Riboulet, T. Roncalli (2000). Copulas for Finance: A Reading Guide and Some Applications. *Working Paper*.

Casella, G., Berger, R. (2002). *Statistical Inference.* Pacific Grove (USA): Duxbury, 2nd edition.

Chai, W-Y., Y. Zhu, Z-Q. Hou (2008). The Research of Copula Theory in Financial Risk Management. *Proceedings of the Seventh International Conference on Machine Learning and Cybernetics*, Kunming, China, 12-15 July.

Charpentier, A., J-D. Fermanian, O. Scaillet. (2007). The Estimation of Copulas: Theory and Practice. In: Rank, Jörn (ed.) "*Copulas: From Theory to Applications in Finance*". London: Risk Books, pp. 35-61.

Cherubini, U., E. Luciano, W. Vecchiato (2004). *Copula Methods in Finance*. West Sussex: John Wiley & Sons Inc.

Conrad, J., G. Kaul (1988). Time-Variation in Expected Returns. *Journal of Business*, Vol. 61, nº 4, pp. 409-425.

Credit Suisse First Boston International - CSFBI (1997). CreditRisk+: A Credit Risk Management Framework. *Technical Report*.

Crook, J., T. Bellotti, (2010). Time Varying and Dynamic Models for Default Risk in Consumer Loans*. Journal of Statistical Society Series A (Statistics in Society)*, Vol. 173, Issue 2, pp. 283-305.

Crook, J., D. Edelman, L. Thomas (2007). Recent developments in consumer credit risk assessment. *European Journal of Operational Research*, 183, 1447-1465.

Crouhy, M., D. Galai, R. Mark (2000). A comparative analysis of current credit risk models. *Journal of Banking & Finance*, 24, pp. 59-117.

Cuadras, C., J. Augé (1981). A continuous general multivariate distribution and its properties. *Comm Statist A – Theory Methods* 10, pp. 339-353.

Czado, Claudia (2010). Pair-copula constructions for multivariate copulas. In: Jaworki, P., F. Durante, W. Härdle, T. Rychlik (eds.) "*Workshop on Copula Theory and its Applications*". Dortrecht: Springer, pp. 93-110.

Das, S., G. Geng (2006). Correlated Default Processes: A Criterion-Based Copula Approach*.* In: Fong, H. (ed.) *“The Credit Market Handbook. Advanced Modeling Issues”*. Hoboken, New Jersey: John Wiley & Sons, pp. 186-218.

Daul, S., E. De Giorgi, F. Lindskog, A. McNeil (2003). The Grouped t-copula with an Application to Credit Risk. *Risk*, 16, pp. 73-76.

De Servigny, A., O. Renault (2004). *Measuring and Managing Credit Risk***.** New York: McGraw Hill.

Di Clemente, A., C. Romano (2004). Measuring and Optimizing Portfolio Credit Risk: A Copula-based Approach*. Economic Notes* (by Banca Monte dei Paschi di Siena SpA), Vol. 33, n^{o} 3, pp. 325-357.

Dubois, H., R. Anderson (2010). Managing households debts: Social service provision in the EU. *Working paper* (European Foundation for the Improvement of Living and Working Conditions).

Durrleman, V., A. Nikeghbali, T. Roncalli (2000). Which copula is the right one? *Working Paper*.

Embrechts, P., F. Lindskog, A. McNeil (2003). Modelling dependence with copulas and applications to risk management. In: Rachev, S. T. (ed.) “*Handbook of Heavy Tailed Distributions in Finance*”. Amsterdam: Elsevier Science, pp. 329-384.

Embrechts, P., A. McNeil, D. Strautman (2002). Correlation and Dependency in Risk Management: Properties and Pitfalls*.* In: Dempster, M.A.H. (ed.) “*Risk Management: Value at Risk and Beyond*”. Cambridge: Cambridge University Press, pp. 176-223.

Engelmann, B., R. Rauhmeier (2006). *The Basel II Risk Parameters. Estimation, Validation, and Stress Testing.* Berlin, Heidelberg: Springer.

Fama, E. F. (1965). The behavior of stock-market prices. *Journal of Business*, 38 (1), pp. 34-105.

Fermanian, J-D., O. Scaillet (2005). Some Statistical Pitfalls in Copula Modeling for Financial Applications. In: Klein, E. (ed.) “*Capital Formation, Governance and Banking*”. New York: Nova Science, pp. 59-74.

Fermanian, J-D., M. Wegkamp (2004). Time Dependent Copulas. *Working Paper*.

Frees, E. W., E. A. Valdez (1998). Understanding Relationships Using Copulas. *North American Actuarial Journal*, Vol. 2, n. 1, pp. 1-25.

Frey, R., A. McNeil (2001). Modelling Dependent Defaults. *Working Paper*.

Frey, R., A. McNeil (2003). Dependent Defaults in Models of Portfolio Credit Risk. *Journal of Risk*, Vol. 6, n^{o} 1.

Frey, R., A. McNeil, M. Nyfeler (2001). Copulas and Credit Models. *Risk*, October, pp. 111-114.

Genest, C., A-C. Favre (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering* - July/August, pp. 347-368.

Genest, C., B. Rémillard, D. Beaudoin (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44, pp. 199-213.

Gordy, Michael B. (2003). A risk-factor model foundation for ratings-based bank capital rules*.* *Journal of Financial Intermediation*, 12, pp. 199-232.

Hamerle, A., D. Rösch (2005). Misspecified Copulas in Credit Risk Models: How Good is Gaussian? *Journal of Risk*, 8 (1), pp. 41-58.

Hamilton, D., J. James, N. Webber (2001). Copula Methods and the Analysis of Credit Risk. *Working Paper*.

Heinen, A., A. Valdesogo (2009). Asymmetric CAPM dependence for large dimensions: the Canonical Vine Autoregressive Model. *ECORE Discussion Paper 2009/101*.

Hull, J., A. White (2006). The Perfect Copula. *Working Paper*.

Hull, J., A. White (2004). Valuation of a CDO and an n^{th} to Default CDS Without Monte Carlo Simulation. *Journal of Derivatives*, 12, 2, Winter.

Joe, Harry (1996). Families of m-variate distributions with given margins and m(m-1)/2 bivariate dependence parameters. *IMS Lecture Notes*, Vol. 28, pp. 120-141.

Joe, Harry (1997). *Multivariate Models and Dependence Concepts.* London, New York: Chapman & Hall/CRC.

Joe, H., J. J. Xu (1996). The estimation method of inference functions for margins for multivariate models. Department of Statistics, University of British Columbia, *Technical Report*, n^{o} 166.

Joe, H., H. Li, A. Nikoloulopoulos (2010). Tail Dependence Functions and Vine Copulas. *Journal of Multivariate Analysis*, Vol. 101, Issue 1, pp. 252-270.

Jouanin, J-F., G. Riboulet, T. Roncalli (2004). Financial Applications of Copula Functions. In: Szegö, G. P. (ed.) "*Risk Measures for the 21st Century*". West Sussex, England: John Wiley & Sons, pp. 273-302.

Kalyvas, L., I. Akkizdis, I. Zourka, V. Bouchereau (2006). *Integrating Market, Credit and Operational Risk. A complete guide for bankers and risk professionals*. Pamplona, Spain: Risk Books.

Kang, W., P. Shahabuddin (2005). Fast simulation for multifactor portfolio credit risk in the t-copula model. *Proceedings of the 2005 Winter Simulation Conference*, pp. 1859-1868.

Kole, E., K. Koedijk, M. Verbeek (2007). Selecting copulas for risk management. *Journal of Banking and Finance*, 31, pp. 2405-2423.

Kolev, N., U. Anjos, B. Mendes (2006). Copulas: A Review and Recent Developments. *Stochastic Models*, 22, pp. 617-660.

Kostadinov, Krassimir (2005). Non-parametric Estimation of Elliptical Copulae With Application to Credit Risk. *Working Paper*.

Kostadinov, Krassimir (2005). Tail approximation for credit risk portfolios with heavy-tailed risk factors. *Technical Report*, The Munich University of Technology.

Kurowicka, D., R. Cooke (2002). The vine copula method for representing high dimensional dependent distributions: application to continuous belief nets. *Proceedings of the 2002 Winter Simulation Conference*.

Kurowicka, D., H. Joe (eds.) (2010). *Dependence Modeling: Handbook on Vine Copulae*. World Scientific.

Lehmann, E. L. (1966). Some Concepts of Dependence. *Annals of Mathematical Statistics*, 37, pp. 1137-1153.

Li, David X. (2000). *On* Default Correlation: A Copula Function Approach. *Journal of Fixed Income*, Vol. 9, n. 4, pp. 43-54.

Lindskog, F., A. McNeil (2001). Common Poisson Shock Models: Applications to Insurance and Credit Risk Modelling. *Working Paper*.

Lo, A., A. MacKinlay (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*, 1, pp. 41-66.

Mandelbrot, B. B. (1963). The variation of certain speculative prices. *Journal of Business*, 36 (4), pp. 394-419.

Markowitz, Harry (1952). Portfolio Selection*. Journal of Finance*, 7, pp. 77-91.

Marshall, A., I. Olkin (1967). A Generalized Bivariate Exponential Distribution. *Journal of Applied Probability*, Vol. 4, n. 2, pp. 291-302.

Maugis, P., D. Guegan (2010). New prospects on vines. *Insurance Markets and Companies: Analyses and Actuarial Computations*, 1, 1 (online).

McNeil, A. J., R. Frey, P. Embrechts (2005). *Quantitative Risk Management. Concepts, Techniques and Tools*. Princeton, UK: Princeton University Press.

Melchiori, Mario R. (2003). Which Archimedean Copula is the right one? *YieldCurve.com e-Journal* (www.Yieldcurve.com).

Mendes, Beatriz V. M. (2005). Computing Conditional VaR Using Time-varying Copulas. *Revista Brasileira de Finanças*, Vol. 3, n^{o} 2, pp. 249-263.

Merton, Robert C. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *Journal of Finance*, 28, pp. 449-470.

Moreira, Fernando F. (2010). Copula-based formulas to estimate unexpected credit losses (The future of Basel Accords?). *Financial Markets, Institutions and Instruments*, Vol. 19, Issue 5, pp. 381-404.

Morillas, Patricia M. (2005). A method to obtain new copulas from a given one. *Metrika*, 61, pp. 169-184.

Nelsen, Roger B. (2006). *An Introduction to Copulas.* New York: Springer. 2nd edition.

Ning, Cathy (2010). Dependence Structure between the Equity Market and the Foreign Exchange Market – a Copula Approach. *Journal of International Money and Finance*, Vol. 29(5), pp. 743-759.

Owen, D., G. Steck (1962). Moments of Order Statistics from the Equicorrelated Multivariate Normal Distribution. *The Annals of Mathematical Statistics*, Vol. 33, n. 4, pp. 1286-1291.

Palaro, H., L. Hotta (2006). Using Conditional Copula to Estimate Value at Risk. *Journal of Data Science*, 4, pp. 93-115.

Patton, Andrew J. (2002). Applications of Copula Theory in Financial Econometrics. University of California, San Diego. *PhD dissertation*.

Patton, Andrew J. (2006). Modelling Asymmetric Exchange Rate Dependence. *International Economic Review*. Vol. 47, n^{o} 2, pp. 527-556.

Perli, R., W. Nayda (2004). Economic and regulatory capital allocation for revolving retail exposures. *Journal of Banking & Finance*, 28, pp. 789-809.

Puzanova, N., S. Siddiqui, M. Trede (2009). Approximate value-at-risk calculation for heterogeneous loan portfolios: Possible enhancements of the Basel II methodology. *Journal of Financial Stability*, 5, pp. 375-392.

Quesada-Molina, José J. (2003). What are copulas? *Monografias Del Seminario de Matemática García de Galdeano*, 27, pp. 499-506.

Rockinger, M., E. Joundeau (2001). Conditional Dependency of Financial Series: An Application of Copulas. *Working Paper*.

Rosenberg, J. V., T. Schuermann (2006). A general approach to integrated risk management with skewed, fat-tailed risks. *Journal of Financial Economics*, 79, pp. 569-614.

Sabato, Gabriele (2006). Managing Credit Risk for Retail Low-Default Portfolios. *Working Paper*.

Savu, C., M. Trede (2006). Hierarchical Archimedean copulas. *Proceedings of International Conference on High Frequency Finance*, Konstanz, Germany, May.

Schmidt, Thorsten (2003). Credit risk modelling and estimation via elliptical copulae. *Working Paper*.

Schmidt, Thorsten (2007). Coping with Copulas. In: Rank, Jörn (ed.) "*Copulas: From Theory to Applications in Finance*". London: Risk Books, pp. 3-34.

Schönbucher, Philipp J. (2000). Factor Models for Portfolio Credit Risk. *Discussion Paper* 16/2001, Bonn Graduate School of Economics, University of Bonn.

Schönbucher, P., D. Schubert (2001). Copula-Dependent Default Risk in Intensity Models. *Working Paper*.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* 8, pp. 229-231.

Thomas, Lyn (2009). Modelling the credit risk for portfolios of consumer loans: analogies with corporate loan models. *Mathematics and Computers in Simulation*, Vol. 79, Issue 8, pp. 2525-2534.

Thomas, L., R. Oliver, D. Hand (2005). A survey of the issues in consumer credit modeling research. *Journal of the Operational Research* Society, 56, pp. 1006-1015.

Trivedi, P. K., D. M. Zimmer (2005). Copula Modeling: An Introduction for Practitioners. *Foundations and Trends® in Econometrics*, Vol. 1, n° 1, pp. 1-111.

Vasicek, Oldrich A. (1991). Limiting Loan Loss Probability Distribution. *KMV Corporation Report*.

Vasicek, Oldrich A. (2002). The Distribution of Loan Portfolio Value. *Risk*.

Weiss, Gregor N. F. (2009). Are Copula-GoF-Tests of Any Practical Use? Results from a Simulation Study. *Working Paper*.

Whelan, Niall (2004). Sampling from Archimedean copulas. *Quantitative Finance*, Vol. 4, Issue 3, pp. 339-352.

APPENDIX A

Best-fit copulas' parameters
(copulas estimated according to the complete default distributions)

TABLE A.1 Copula parameters estimated for pairs AB, AC, AD and AE (best-fit based on complete default distributions)

GoF APPROACH	AB		AC		AD		AE	
	Copula	Parameter	Copula	Parameter	Copula	Parameter	Copula	Parameter
Empirical Copula	Clayton	2.9580448	Clayton	3.0354896	Galambos	0.0125000	Student t	-0.4430353
Kendall's Transform	Plackett	7.6454102	Plackett	8.2162109	Galambos	0.0125000	Frank	-3.3524919
Rosenblatt's Transform	Frank	4.7446299	Clayton	3.0354896	Gumbel	1.0000014	Plackett	0.2063477

TABLE A.2 Copula parameters estimated for pairs BC, BD and BE (best-fit based on complete default distributions)

GoF APPROACH	BC		BD		BE	
	Copula	Parameter	Copula	Parameter	Copula	Parameter
Empirical Copula	Clayton	11.3340991	Hüsler-Reiss	0.6412109	Student t	-0.5058982
Kendall's Transform	Clayton	11.3340991	Galambos	0.0125000	Plackett	0.1311523
Rosenblatt's Transform	Plackett	161.3443359	Joe	1.1007813	Plackett	0.1311523

TABLE A.3 Copula parameters estimated for pairs CD, CE and DE (best-fit based on complete default distributions)

GoF APPROACH	CD		CE		DE	
	Copula	Parameter	Copula	Parameter	Copula	Parameter
Empirical Copula	Hüsler-Reiss	0.6699219	Gaussian	-0.4534993	Plackett	3.5805664
Kendall's Transform	Joe	1.1268555	Frank	-4.5301524	FGM	0.7535625
Rosenblatt's Transform	Joe	1.1268555	Plackett	0.0832031	Plackett	3.5805664

APPENDIX B

Detailed results of goodness-of-fit tests based on the complete default rate distributions

In the following tables, "distance" is the measure of the difference between the (whole) empirical distribution of default rates and the candidate copulas. The p-values are related to the null hypothesis that the underlying copula C belongs to the family of candidate copula C_0 or some transformation of the copula C belongs to the family of transformations of that copula, that is, H_0: C $\in$ C_0 or H_0: $C^T \in C^T_0$. So, in order to confirm the significance of the candidate copula high p-values are expected.

TABLE B.1: Detailed results of copula estimation considering three GoF approaches (pair AB) and complete default rate distributions

METHOD / COPULA	Empirical-copula		Kendall's Transform		Rosenblatt's Transform	
	"Distance"	p-value	"Distance"	p-value	"Distance"	p-value
Gaussian	0.0762	0.0050	0.0809	0.0330	0.0709	0.0340
t	0.0763	0.0040	0.0895	0.0150	0.0710	0.0200
Clayton	**0.0599**	**0.0230**	0.0930	0.0130	0.0953	0.0190
Frank	0.0824	0.0030	0.0809	0.0080	**0.0509**	0.5900
Gumbel	0.1197	0.0010	0.1205	0.0020	0.1960	0.0710
FGM	-	0.0000	-	0.0000	-	0.0000
Galambos	0.2224	0.0000	0.2393	0.0010	0.1001	0.0220
Hüsler-Reiss	0.1256	0.0050	0.1291	0.0050	0.0680	0.0150
Joe	0.1774	0.0150	0.1911	0.0040	0.0802	0.0420
Plackett	0.0789	0.0050	**0.0740**	**0.0190**	0.0609	**0.0800**

Note: the smallest distance for each method is highlighted in boldface. The symbol " - " means that the estimations did not yield values compatible with the parameter domain for the respective copula. In such cases, the p-value was set equal to zero (which implies the rejection of the candidate copula).

TABLE B.2: Detailed results of copula estimation considering three GoF approaches (pair AC) and complete default rate distributions

METHOD / COPULA	Empirical-copula		Kendall's Transform		Rosenblatt's Transform	
	"Distance"	**p-value**	**"Distance"**	**p-value**	**"Distance"**	**p-value**
Gaussian	0.05578	0.04196	0.06370	0.08392	0.06555	0.05200
t	0.05584	0.02797	0.06370	0.04895	0.06577	0.05100
Clayton	**0.04815**	**0.05495**	0.08609	0.01698	**0.04467**	**0.45200**
Frank	0.06348	0.00899	0.09297	0.00200	0.04981	0.57600
Gumbel	0.09237	0.00999	0.09986	0.01399	0.22183	0.06400
FGM	-	0.00000	-	0.00000	-	0.00000
Galambos	0.20935	0.00000	0.21694	0.00200	0.11115	0.01600
Hüsler-Reiss	0.10218	0.00999	0.11019	0.00500	0.07504	0.00400
Joe	0.15125	0.02697	0.16529	0.00599	0.08337	0.02900
Plackett	0.05900	0.01499	**0.05682**	**0.05395**	0.05141	0.17800

Note: the smallest distance for each method is highlighted in boldface. The symbol " - " means that the estimations did not yield values compatible with the parameter domain for the respective copula. In such cases, the p-value was set equal to zero (which implies the rejection of the candidate copula).

TABLE B.3: Detailed results of copula estimation considering three GoF approaches (pair AD) and complete default rate distributions

METHOD / COPULA	Empirical-copula		Kendall's Transform		Rosenblatt's Transform	
	"Distance"	**p-value**	**"Distance"**	**p-value**	**"Distance"**	**p-value**
Gaussian	0.02191	0.46254	0.04304	0.20879	0.02038	0.99500
t	0.02206	0.40859	0.04477	0.18681	0.02065	0.99900
Clayton	0.04053	0.14585	0.09986	0.03696	0.03080	0.77300
Frank	0.02198	0.34166	0.04304	0.15185	0.02403	0.93700
Gumbel	0.02173	0.56843	0.03788	0.47353	**0.01989**	**0.96200**
FGM	0.02197	0.70829	0.04304	0.43656	0.02048	0.95900
Galambos	**0.02173**	**0.64635**	**0.03788**	**0.47852**	0.01989	0.92100
Hüsler-Reiss	0.02173	0.61439	0.03788	0.47253	0.01989	0.91100
Joe	0.14484	0.05395	0.11708	0.03097	0.11614	0.00600
Plackett	0.02199	0.32567	0.04304	0.15185	0.02052	1.00000

Note: the smallest distance for each method is highlighted in boldface. In cases of tie, the distance presenting the highest p-value (lowest probability of rejecting the candidate copula) was selected.

TABLE B.4: Detailed results of copula estimation considering three GoF approaches (pair AE) and complete default rate distributions

METHOD / COPULA	Empirical-copula		Kendall's Transform		Rosenblatt's Transform	
	"Distance"	p-value	"Distance"	p-value	"Distance"	p-value
Gaussian	0.0164	0.4056	0.0155	0.7173	0.0563	0.3140
t	**0.0160**	**0.3856**	0.0155	0.6923	0.0571	0.3680
Clayton	0.0643	0.0599	0.1085	0.0779	0.0728	0.1000
Frank	0.0195	0.1788	**0.0121**	**0.6753**	2.2273	0.6560
Gumbel	0.0643	0.0669	0.1085	0.0579	0.0728	0.1190
FGM	0.0193	0.6643	0.0258	0.7113	0.0550	0.4640
Galambos	0.0643	0.1219	0.1085	0.0639	0.0728	0.0870
Hüsler-Reiss	0.0643	0.0929	0.1085	0.0519	0.0728	0.0770
Joe	0.0632	0.7303	0.0224	0.9840	0.0987	0.0200
Plackett	0.0189	0.2707	0.0155	0.6513	**0.0502**	**0.4880**

Note: the smallest distance for each method is highlighted in boldface.

TABLE B.5: Detailed results of copula estimation considering three GoF approaches (pair BC) and complete default rate distributions

METHOD / COPULA	Empirical-copula		Kendall's Transform		Rosenblatt's Transform	
	"Distance"	p-value	"Distance"	p-value	"Distance"	p-value
Gaussian	0.01302	0.71329	0.02755	0.36963	0.08392	0.12000
t	0.01330	0.47952	0.02927	0.29171	0.04170	0.80100
Clayton	**0.01016**	**0.77522**	**0.01894**	**0.68631**	0.04057	0.90600
Frank	0.01398	0.58042	0.03443	0.36364	0.04400	0.62600
Gumbel	0.01744	0.48551	0.02583	0.47752	0.81997	0.00000
FGM	-	0.00000	-	0.00000	-	0.00000
Galambos	0.02533	0.99600	0.03443	0.98901	0.17828	0.78700
Hüsler-Reiss	0.02244	0.97003	0.02066	0.98002	0.14082	0.00100
Joe	0.04370	0.73227	0.02927	0.75724	0.10734	0.11000
Plackett	0.01283	0.60939	0.03443	0.16184	**0.03795**	**0.80400**

Note: the smallest distance for each method is highlighted in boldface. The symbol " - " means that the estimations did not yield values compatible with the parameter domain for the respective copula. In such cases, the p-value was set equal to zero (which implies the rejection of the candidate copula).

TABLE B.6: Detailed results of copula estimation considering three GoF approaches (pair BD) and complete default rate distributions

METHOD / COPULA	Empirical-copula		Kendall's Transform		Rosenblatt's Transform	
	"Distance"	p-value	"Distance"	p-value	"Distance"	p-value
Gaussian	0.02506	0.38462	0.04821	0.16484	0.02645	0.84900
t	0.02267	0.44256	0.04304	0.24276	0.02676	0.85100
Clayton	0.03863	0.19680	0.08092	0.10989	0.04355	0.36500
Frank	0.02418	0.32368	0.04477	0.15485	0.04646	0.78200
Gumbel	0.02184	0.60939	0.03443	0.41758	0.03566	0.49700
FGM	0.02474	0.67532	0.04649	0.37263	0.02568	0.81100
Galambos	0.02781	0.50350	**0.02583**	**0.68032**	0.02311	0.82300
Hüsler-Reiss	**0.02172**	**0.60440**	0.03271	0.47552	0.02064	0.93500
Joe	0.02346	0.68931	0.02583	0.66833	**0.02012**	**0.91600**
Plackett	0.02509	0.29071	0.04649	0.14885	0.02666	0.84600

Note: the smallest distance for each method is highlighted in boldface. In cases of tie, the distance presenting the highest p-value (lowest probability of rejecting the candidate copula) was selected.

TABLE B.7: Detailed results of copula estimation considering three GoF approaches (pair BE) and complete default rate distributions

METHOD / COPULA	Empirical-copula		Kendall's Transform		Rosenblatt's Transform	
	"Distance"	p-value	"Distance"	p-value	"Distance"	p-value
Gaussian	0.01394	0.53946	0.02755	0.31269	0.06638	0.17500
t	**0.01027**	**0.64935**	0.02066	0.42358	0.06159	0.42700
Clayton	0.08073	0.02697	0.15496	0.01499	0.08460	0.05500
Frank	0.01257	0.35664	0.00689	0.88911	2.36373	0.78900
Gumbel	0.08073	0.03097	0.15496	0.01099	0.08460	0.06500
FGM	0.02157	0.63836	0.04477	0.44555	0.06740	0.30700
Galambos	0.08073	0.06693	0.15496	0.02098	0.08460	0.04700
Hüsler-Reiss	0.08073	0.06394	0.15496	0.01698	0.08460	0.03500
Joe	0.07339	0.52448	0.04477	0.52947	0.11346	0.01000
Plackett	0.01370	0.41558	**0.00689**	**0.92507**	**0.04615**	**0.70100**

Note: the smallest distance for each method is highlighted in boldface. In cases of tie, the distance presenting the highest p-value (lowest probability of rejecting the candidate copula) was selected.

TABLE B.8: Detailed results of copula estimation considering three GoF approaches (pair CD) and complete default rate distributions

METHOD / COPULA	Empirical-copula		Kendall's Transform		Rosenblatt's Transform	
	"Distance"	p-value	"Distance"	p-value	"Distance"	p-value
Gaussian	0.02511	0.41558	0.04993	0.18382	0.02618	0.85100
t	0.02319	0.41758	0.05165	0.14386	0.02742	0.81800
Clayton	0.04103	0.16184	0.08436	0.08691	0.05330	0.16900
Frank	0.02568	0.29471	0.04821	0.12388	0.04041	0.87900
Gumbel	0.02069	0.65934	0.03443	0.41958	0.04667	0.35500
FGM	0.02585	0.64935	0.04821	0.34066	0.02634	0.77700
Galambos	0.02847	0.46853	0.03099	0.59540	0.02462	0.77100
Hüsler-Reiss	**0.02036**	**0.63337**	0.02755	0.61439	0.02017	0.94900
Joe	0.02204	0.71728	**0.02755**	**0.64036**	**0.01979**	**0.94600**
Plackett	0.02711	0.26573	0.04993	0.09590	0.02857	0.79700

Note: the smallest distance for each method is highlighted in boldface. In cases of tie, the distance presenting the highest p-value (lowest probability of rejecting the candidate copula) was selected.

TABLE B.9: Detailed results of copula estimation considering three GoF approaches (pair CE) and complete default rate distributions

METHOD / COPULA	Empirical-copula		Kendall's Transform		Rosenblatt's Transform	
	"Distance"	p-value	"Distance"	p-value	"Distance"	p-value
Gaussian	**0.00982**	**0.73127**	0.01205	0.82517	0.07110	0.14300
t	0.01125	0.57942	0.00861	0.87712	0.06520	0.51400
Clayton	0.05679	0.08292	0.12052	0.04695	0.08913	0.04300
Frank	0.02296	0.06194	**0.00344**	**0.98002**	2.48079	0.81400
Gumbel	0.05679	0.07792	0.12052	0.02697	0.08913	0.03200
FGM	0.01260	0.86414	0.02755	0.70230	0.07359	0.30400
Galambos	0.05679	0.16184	0.12052	0.04595	0.08913	0.04300
Hüsler-Reiss	0.05679	0.14186	0.12052	0.03896	0.08913	0.05000
Joe	0.08408	0.36563	0.04132	0.60939	0.11932	0.00600
Plackett	0.03286	0.03497	0.01033	0.80919	**0.05388**	**0.62700**

Note: the smallest distance for each method is highlighted in boldface.

TABLE B.10: Detailed results of copula estimation considering three GoF approaches (pair DE) and complete default rate distributions

METHOD / COPULA	Empirical-copula		Kendall's Transform		Rosenblatt's Transform	
	"Distance"	p-value	"Distance"	p-value	"Distance"	p-value
Gaussian	0.04082	0.13487	0.03788	0.37063	0.03982	0.28300
t	0.02298	0.44855	0.02066	0.82717	0.03488	0.50800
Clayton	0.01626	0.78322	0.03443	0.70330	0.04008	0.50800
Frank	0.01750	0.60639	0.02066	0.82517	0.15382	0.00800
Gumbel	0.05927	0.06494	0.03788	0.38861	0.07205	0.12900
FGM	0.03094	0.60040	**0.01722**	**0.92507**	0.03651	0.48500
Galambos	0.09137	0.03896	0.06543	0.21479	0.05012	0.23000
Hüsler-Reiss	0.09137	0.04096	0.06543	0.18482	0.05012	0.21300
Joe	0.11223	0.19880	0.08609	0.14985	0.06134	0.26900
Plackett	**0.01622**	**0.69131**	0.02066	0.85514	**0.03190**	**0.63800**

Note: the smallest distance for each method is highlighted in boldface.

APPENDIX C

Best-fit copulas' parameters
(copulas estimated according to the right tails of the default distributions)

TABLE C.1 Copula parameters estimated for pairs AB, AC, AD and AE (best-fit based on the right tails of default distributions)

GoF APPROACH	AB		AC		AD		AE	
	Copula	**Parameter**	**Copula**	**Parameter**	**Copula**	**Parameter**	**Copula**	**Parameter**
Empirical Copula	Galambos	0.0125000	Joe	1.2684570	Galambos	0.0125000	Frank	-3.3524919
Kendall's Transform	Galambos	0.0125000	Galambos	0.0125000	Gumbel	1.0000014	Clayton	0.0000015
Rosenblatt's Transform	Frank	4.7446299	Clayton	3.0354896	Gumbel	1.0000014	Plackett	0.2063477

TABLE C.2 Copula parameters estimated for pairs BC, BD and BE (best-fit based on the right tails of default distributions)

GoF APPROACH	BC		BD		BE	
	Copula	**Parameter**	**Copula**	**Parameter**	**Copula**	**Parameter**
Empirical Copula	Clayton	11.3340991	Gumbel	1.1175104	Frank	-3.9901586
Kendall's Transform	Student t	0.9622511	Clayton	0.7649150	Clayton	0.0000015
Rosenblatt's Transform	Plackett	161.3443359	Joe	1.1007813	Plackett	0.1311523

TABLE C.3 Copula parameters estimated for pairs CD, CE and DE (best-fit based on the right tails of default distributions)

GoF APPROACH	CD		CE		DE	
	Copula	**Parameter**	**Copula**	**Parameter**	**Copula**	**Parameter**
Empirical Copula	Joe	1.1268555	Plackett	0.0832031	Hüsler-Reiss	0.0999999
Kendall's Transform	Clayton	0.8203320	Clayton	0.0000015	Hüsler-Reiss	0.0999999
Rosenblatt's Transform	Joe	1.1268555	Plackett	0.0832031	Plackett	3.5805664

APPENDIX D

Detailed results of goodness-of-fit tests based on the best-fit to the right tails

In the following tables, “distance” is the measure of the difference between the empirical distribution of default rates and the candidate copulas. The p-values are related to the null hypothesis that the underlying copula C belongs to the family of candidate copula C_0 or some transformation of the copula C belongs to the family of transformations of that copula, that is, H_0: $C \in C_0$ or H_0: $C^T \in C^T{}_0$. So, in order to confirm the significance of the candidate copula high p-values are expected.

TABLE D.1: Detailed results of copula estimation considering three GoF methods (pair AB) and the right tail of the default rate distributions

METHOD / COPULA	Empirical-copula		Kendall’s Transform		Rosenblatt’s Transform	
	“Distance”	p-value	“Distance”	p-value	“Distance”	p-value
Gaussian	0.00622	0.11988	0.08333	0.00400	0.07094	0.02800
t	0.00615	0.10789	0.00344	0.01399	0.07099	0.02400
Clayton	0.00551	0.28671	0.02778	0.57343	0.09533	0.02100
Frank	0.00564	0.19980	0.08333	0.00599	**0.05091**	**0.58600**
Gumbel	0.00416	0.23776	0.08333	0.00400	0.19605	0.05600
FGM	-	0.00000	-	0.00000	-	0.00000
Galambos	**0.00116**	**0.61139**	**0.00000**	**0.17882**	0.10009	0.02500
Hüsler-Reiss	0.00339	0.22078	0.05556	0.06693	0.06803	0.01600
Joe	0.00129	0.85015	0.05556	0.07493	0.08017	0.02900
Plackett	0.00637	0.14685	0.08333	0.00599	0.06090	0.09600

Note: the smallest distance for each method is highlighted in boldface. The symbol “ - ” means that the estimations did not yield values compatible with the parameter domain for the respective copula. In such cases, the p-value was set equal to zero (which implies the rejection of the candidate copula).

TABLE D.2: Detailed results of copula estimation considering three GoF methods (pair AC) and the right tail of the default rate distributions

METHOD / COPULA	Empirical-copula		Kendall's Transform		Rosenblatt's Transform	
	"Distance"	p-value	"Distance"	p-value	"Distance"	p-value
Gaussian	0.00328	0.41558	0.08333	0.00400	0.06555	0.05000
t	0.00323	0.41459	0.00344	0.00999	0.06577	0.04400
Clayton	0.00202	0.69231	0.02778	0.57143	**0.04467**	**0.43000**
Frank	0.00257	0.58042	0.08333	0.01099	0.04981	0.58100
Gumbel	0.00242	0.51948	0.08333	0.00300	0.22183	0.06600
FGM	-	0.00000	-	0.00000	-	0.00000
Galambos	0.00295	0.18082	**0.00000**	**0.18182**	0.11115	0.00800
Hüsler-Reiss	0.00159	0.61638	0.08333	0.00100	0.07504	0.01200
Joe	**0.00086**	**0.94006**	0.05556	0.08991	0.08337	0.03200
Plackett	0.00299	0.49051	0.08333	0.00300	0.05141	0.20400

Note: the smallest distance for each method is highlighted in boldface. The symbol " - " means that the estimations did not yield values compatible with the parameter domain for the respective copula. In such cases, the p-value was set equal to zero (which implies the rejection of the candidate copula).

TABLE D.3: Detailed results of copula estimation considering three GoF methods (pair AD) and the right tail of the default rate distributions

METHOD / COPULA	Empirical-copula		Kendall's Transform		Rosenblatt's Transform	
	"Distance"	p-value	"Distance"	p-value	"Distance"	p-value
Gaussian	0.00100	0.65435	0.02778	0.09291	0.02038	0.98900
t	0.00105	0.66134	0.00344	0.00000	0.02065	0.99700
Clayton	0.00193	0.61638	0.02778	0.10789	0.03080	0.79000
Frank	0.00094	0.70030	0.02778	0.06993	0.02403	0.92300
Gumbel	0.00080	0.71828	**0.00000**	**0.27073**	**0.01989**	**0.93400**
FGM	0.00093	0.74426	0.02778	0.06893	0.02048	0.96500
Galambos	**0.00080**	**0.73127**	0.00000	0.17782	0.01989	0.93700
Hüsler-Reiss	0.00080	0.68531	0.00000	0.14985	0.01989	0.90600
Joe	0.03836	0.65834	0.05556	0.00699	0.11614	0.00600
Plackett	0.00094	0.71828	0.02778	0.07493	0.02052	1.00000

Note: the smallest distance for each method is highlighted in boldface. In cases of tie, the distance presenting the highest p-value (lowest probability of rejecting the candidate copula) was selected.

TABLE D.4: Detailed results of copula estimation considering three GoF methods (pair AE) and the right tail of the default rate distributions

METHOD / COPULA	Empirical-copula		Kendall's Transform		Rosenblatt's Transform	
	"Distance"	p-value	"Distance"	p-value	"Distance"	p-value
Gaussian	0.0002	0.6154	0.0000	0.0430	0.0563	0.3030
t	0.0003	0.5405	0.0017	0.0819	0.0571	0.3630
Clayton	0.0025	0.3077	**0.0000**	**0.4685**	0.0728	0.0970
Frank	**0.0001**	**0.5694**	0.0000	0.0350	2.2273	0.6870
Gumbel	0.0025	0.1668	0.0000	0.2458	0.0728	0.1150
FGM	0.0004	0.8871	0.0000	0.1728	0.0550	0.4520
Galambos	0.0025	0.2857	0.0000	0.1838	0.0728	0.0840
Hüsler-Reiss	0.0025	0.2777	0.0000	0.1538	0.0728	0.0840
Joe	0.0424	0.3157	0.0556	0.0040	0.0987	0.0110
Plackett	0.0001	0.6354	0.0000	0.0370	**0.0502**	**0.5110**

Note: the smallest distance for each method is highlighted in boldface. In cases of tie, the distance presenting the highest p-value (lowest probability of rejecting the candidate copula) was selected.

TABLE D.5: Detailed results of copula estimation considering three GoF methods (pair BC) and the right tail of the default rate distributions

METHOD / COPULA	Empirical-copula		Kendall's Transform		Rosenblatt's Transform	
	"Distance"	p-value	"Distance"	p-value	"Distance"	p-value
Gaussian	0.00388	0.19281	0.05556	0.35365	0.08392	0.11200
t	0.00460	0.07193	**0.00172**	**0.02198**	0.04170	0.81800
Clayton	**0.00383**	**0.40859**	0.02778	0.86014	0.04057	0.87600
Frank	0.00428	0.21578	0.05556	0.54446	0.04400	0.65600
Gumbel	0.00482	0.08591	0.05556	0.25175	0.81997	0.00000
FGM	-	0.00000	-	0.00000	-	0.00000
Galambos	0.03232	0.00000	0.11111	0.00000	0.23423	0.00000
Hüsler-Reiss	0.00389	0.09491	0.05556	0.04396	0.14082	0.00200
Joe	0.00476	0.56543	0.05556	0.68332	0.10734	0.10600
Plackett	0.00421	0.16084	0.05556	0.47552	**0.03795**	**0.77400**

Note: the smallest distance for each method is highlighted in boldface. The symbol " - " means that the estimations did not yield values compatible with the parameter domain for the respective copula. In such cases, the p-value was set equal to zero (which implies the rejection of the candidate copula).

TABLE D.6: Detailed results of copula estimation considering three GoF methods (pair BD) and the right tail of the default rate distributions

METHOD / COPULA	Empirical-copula		Kendall's Transform		Rosenblatt's Transform	
	"Distance"	**p-value**	**"Distance"**	**p-value**	**"Distance"**	**p-value**
Gaussian	0.00171	0.60839	0.00000	0.76623	0.02645	0.86400
t	0.00158	0.61938	0.00000	0.70729	0.02676	0.85800
Clayton	0.00197	0.58641	**0.00000**	**0.85614**	0.04355	0.35800
Frank	0.00187	0.55345	0.00000	0.75225	0.04646	0.78100
Gumbel	**0.00139**	**0.67133**	0.02778	0.15884	0.03566	0.48400
FGM	0.00192	0.63037	0.00000	0.74326	0.02568	0.81500
Galambos	0.00260	0.23576	0.02778	0.07193	0.02311	0.82900
Hüsler-Reiss	0.00143	0.61439	0.00000	0.51848	0.02064	0.93300
Joe	0.00144	0.79321	0.00000	0.57343	**0.02012**	**0.91600**
Plackett	0.00185	0.56743	0.00000	0.74625	0.02666	0.88100

Note: the smallest distance for each method is highlighted in boldface. In cases of tie, the distance presenting the highest p-value (lowest probability of rejecting the candidate copula) was selected.

TABLE D.7: Detailed results of copula estimation considering three GoF methods (pair BE) and the right tail of the default rate distributions

METHOD / COPULA	Empirical-copula		Kendall's Transform		Rosenblatt's Transform	
	"Distance"	**p-value**	**"Distance"**	**p-value**	**"Distance"**	**p-value**
Gaussian	0.00015	0.59740	0.00000	0.04096	0.06638	0.19600
t	0.00023	0.47752	0.00172	0.06993	0.06159	0.40500
Clayton	0.00245	0.32867	**0.00000**	**0.48052**	0.08460	0.05200
Frank	**0.00006**	**0.75025**	0.00000	0.01998	2.36373	0.77300
Gumbel	0.00245	0.19381	0.00000	0.25774	0.08460	0.06600
FGM	0.00039	0.91209	0.00000	0.19281	0.06740	0.35500
Galambos	0.00245	0.26873	0.00000	0.16084	0.08460	0.05700
Hüsler-Reiss	0.00245	0.28172	0.00000	0.16783	0.08460	0.04600
Joe	0.04251	0.30070	0.05556	0.00300	0.11346	0.00400
Plackett	0.00007	0.72228	0.00000	0.03497	**0.04615**	**0.68800**

Note: the smallest distance for each method is highlighted in boldface. In cases of tie, the distance presenting the highest p-value (lowest probability of rejecting the candidate copula) was selected.

TABLE D.8: Detailed results of copula estimation considering three GoF methods (pair CD) and the right tail of the default rate distributions

METHOD / COPULA	Empirical-copula		Kendall's Transform		Rosenblatt's Transform	
	"Distance"	p-value	"Distance"	p-value	"Distance"	p-value
Gaussian	0.00178	0.61439	0.00000	0.78022	0.02618	0.88300
t	0.00172	0.62338	0.00000	0.70729	0.02742	0.82500
Clayton	0.00199	0.62537	**0.00000**	**0.83816**	0.05330	0.16000
Frank	0.00189	0.53546	0.00000	0.77023	0.04041	0.87000
Gumbel	0.00143	0.66434	0.02778	0.20180	0.04667	0.33300
FGM	0.00193	0.63137	0.00000	0.76923	0.02634	0.76900
Galambos	0.00260	0.21778	0.02778	0.06394	0.02462	0.79800
Hüsler-Reiss	0.00139	0.62138	0.00000	0.53147	0.02017	0.94400
Joe	**0.00133**	**0.83417**	0.02778	0.27373	**0.01979**	**0.93500**
Plackett	0.00190	0.55145	0.00000	0.77722	0.02857	0.77300

Note: the smallest distance for each method is highlighted in boldface. In cases of tie, the distance presenting the highest p-value (lowest probability of rejecting the candidate copula) was selected.

TABLE D.9: Detailed results of copula estimation considering three GoF methods (pair CE) and the right tail of the default rate distributions

METHOD / COPULA	Empirical-copula		Kendall's Transform		Rosenblatt's Transform	
	"Distance"	p-value	"Distance"	p-value	"Distance"	p-value
Gaussian	0.00013	0.60939	0.00000	0.02897	0.07110	0.15400
t	0.00024	0.41059	0.00172	0.08392	0.06520	0.52600
Clayton	0.00245	0.34765	**0.00000**	**0.51548**	0.08913	0.04500
Frank	0.00006	0.98102	0.00000	0.01598	2.48079	0.83300
Gumbel	0.00245	0.19880	0.00000	0.28571	0.08913	0.06100
FGM	0.00040	0.92507	0.00000	0.17782	0.07359	0.26800
Galambos	0.00245	0.31868	0.00000	0.19780	0.08913	0.03400
Hüsler-Reiss	0.00245	0.28472	0.00000	0.16384	0.08913	0.03300
Joe	0.04302	0.25075	0.05556	0.00500	0.11932	0.00400
Plackett	**0.00005**	**0.97902**	0.00000	0.02398	**0.05388**	**0.63400**

Note: the smallest distance for each method is highlighted in boldface. In cases of tie, the distance presenting the highest p-value (lowest probability of rejecting the candidate copula) was selected.

TABLE D.10: Detailed results of copula estimation considering three GoF methods (pair DE) and the right tail of the default rate distributions

METHOD / COPULA	Empirical-copula		Kendall's Transform		Rosenblatt's Transform	
	"Distance"	**p-value**	**"Distance"**	**p-value**	**"Distance"**	**p-value**
Gaussian	0.00265	0.37163	0.02778	0.22977	0.03982	0.29200
t	0.00691	0.07992	0.00344	0.00400	0.03488	0.48300
Clayton	0.00314	0.42358	0.02778	0.15485	0.04008	0.49400
Frank	0.00469	0.21778	0.02778	0.22677	0.03728	0.93400
Gumbel	0.00263	0.27473	0.02778	0.26074	0.07205	0.11800
FGM	0.00255	0.47652	0.02778	0.11489	0.03651	0.48100
Galambos	0.00080	0.72128	0.00000	0.17682	0.05012	0.22400
Hüsler-Reiss	0.00080	0.74625	0.00000	0.17682	0.05012	0.23100
Joe	**0.00045**	**0.99700**	**0.00000**	**0.64535**	0.06134	0.27100
Plackett	0.00613	0.11189	0.02778	0.33067	**0.03190**	**0.65600**

Note: the smallest distance for each method is highlighted in boldface. In cases of tie, the distance presenting the highest p-value (lowest probability of rejecting the candidate copula) was selected. Although the Joe copula is the best-fit according to the Empirical-copula and the Kendall's Transform approaches, its parameter estimated via Canonical Maximum Likelihood (0.9216797) is smaller than the minimum (1) allowed for that family. So the second best copula, Hüsler-Reiss, is considered the most representative family for this pair based on those two approaches.

APPENDIX E

Derivation of a trivariate density function – expression [4.5] (based on Aas et al., 2009)

In the derivation below, $f(.)$ is the density function, d is the number of variables, and $c(.)$ is the density copula for the associate pair.

From $f(x_1,..., x_d) = f_d(x_d).f(x_{d-1} \mid x_d).f(x_{d-2} \mid x_{d-1}, x_d)...f(x_1 \mid x_2,..., x_d)$ presented in Section 4.2.1. Selecting $d = 3$ we have:

$$f(x_1, x_2, x_3) = f_3(x_3).f(x_2 \mid x_3).f(x_1 \mid x_2, x_3) \quad \text{[E.1]}$$

Now decomposing the two last terms above:

(i) $f(x_2 \mid x_3) = c_{23}(F_2(x_2), F_3(x_3)).f_2(x_2)$

(ii.1) $f(x_1 \mid x_2, x_3) = f(x_1, x_2 \mid x_3) / f(x_2 \mid x_3)$; then decomposing again:

(ii.2) $f(x_1, x_2 \mid x_3) = c_{12|3}(F(x_1 \mid x_3), F(x_2 \mid x_3)).f(x_1 \mid x_3).f(x_2 \mid x_3)$

Plugging (ii.2) into (ii.1):

(ii.3) $f(x_1 \mid x_2, x_3) = c_{12|3}(F(x_1 \mid x_3), F(x_2 \mid x_3)).f(x_1 \mid x_3)$

Decomposing the last conditional density in (ii.3):

(ii.4) $f(x_1 \mid x_3) = c_{13}(F_1(x_1), F_3(x_3)).f_1(x_1)$

An extended expression for (ii.3) is:

(ii.5) $f(x_1 \mid x_2, x_3) = c_{12|3}(F(x_1 \mid x_3), F(x_2 \mid x_3)).c_{13}(F_1(x_1), F_3(x_3)).f_1(x_1)$

Finally, plugging (i) and (ii.5) into [E.1], we get:

$$f(x_1, x_2, x_3) =$$

$$f_3(x_3).c_{23}(F_2(x_2), F_3(x_3)).f_2(x_2).c_{12|3}(F(x_1 \mid x_3), F(x_2 \mid x_3)).c_{13}(F_1(x_1), F_3(x_3)).f_1(x_1) =$$

$$f_1(x_1).f_2(x_2).f_3(x_3).c_{13}(F_1(x_1), F_3(x_3)).c_{23}(F_2(x_2), F_3(x_3)).c_{12|3}(F(x_1 \mid x_3), F(x_2 \mid x_3))$$

which is formula [4.5] in the text.