**Default Dependency in Peer-to-Peer Lending: A Copula Approach**

**Leyla Mammadova**

**Abstract**

We explore default dependency in P2P lending with a specific focus on asymmetric tail dependency. We use five different copula families to capture possible non-linearity and asymmetric properties of the loan portfolios. We apply monthly dataset of “Funding Circle” from 08/2010 to 10/2018 by covering six different risky loan segments and therefore, fifteen loan portfolios pairs. From the estimations of entire distribution and upper tails of fifteen portfolio pairs, we find that the Gumbel copula representing right tail dependence fit the empirical data better others in most cases, followed by the copula in three cases by exhibiting symmetric tail dependence. Our results suggest that not only individual defaults can be a concerning factor in P2P lending, but also default dependency, particularly in the upper tails of joint default distributions. We also compare the normal and copula approaches with the original dataset in predicting joint extreme defaults. The copula methods outperform the traditional normal method 52 out of 75 cases. The normal method, however, produce closer estimation in least risky and highest risky segments. Moreover, the findings do not support the hypothesis that normal method underestimate extreme defaults. Conversely, we find that in most cases, the normal method estimates considerably higher joint default rates than the original dataset and the copula method.

1. **Introduction**

Recent survey among the UK platforms reveals increased concern of participants toward default risk. Specifically, 40% of P2P business lending platforms indicate default as “high” or “very high” risk level (Zhang, et al., 2018). Although, individual defaults are not the major concern of credit risk if the portfolio is well diversified, in the case of strong systematic dependency between obligators even a portfolio including many small loans can be highly risky.

Default dependency has been widely explored in the credit risk management literature as it affects the distribution of loan portfolio losses and is therefore critical in determining quantiles or other risk measure used for allocating capital for solvency purposes (Lando, 2004). Frey, McNeil, and Nyfeler (2003) show that joint default events have an important impact on the performance of loans as payoff is linked to the loss of whole portfolio. Due to the contingency effect, the default of one counter party can drive the default of others and large unexpected losses in the case of strong systematic dependency (Servigny, 2004).

Developing default dependency models has been a major concern as part of general development towards an active and market-based management of credit risks in banks and other financial institutions. Different models or small moderations to the models describing dependency can have a considerable impact on the resulting credit loss distribution, particularly in tail. The literature on default correlation, concerning on large portfolios tend to focus on traditional models such as JP Morgan’s Credit Metrics (1997) and Credit Suisse Credit Risk+ (1997). However, this kind of models can be subjected to considerable model risk, as they hold simplified linearity assumptions between individual defaults which are not often the case in credit risk (Frey, McNeil, and Nyfeler, 2001; Frey and McNeil, 2002). Instead, a number studies (Embrechts et al, 2002; Nikoloulopoulos et al, 2012; Fang and Madsen, 2013) apply copula models as they are suitable to capture non-linearity and asymmetric dependence between individual defaults of a portfolio.

However, the similar studies have not been carried out of P2P lending, given it is relatively new phenomenon in financial markets. While P2P lending is considered less risky than traditional banking, its novel nature and lack of experience in the stress scenarios might be the indications to be less confident regarding the risk framework of P2P lending. Peer to peer platforms (P2P) transfers the traditional idea of bank lending to online platforms. The platforms do not take risk on the balance sheet, as they do not accept deposits from customers (Deloitte, 2016). However, investors bear the entire credit risk because there is no intermediary between borrowers and lenders. Apart from individual defaults, joint defaults may create additional systematic risks in the market. Therefore, it is worthwhile to explore default dependency in P2P lending. Moreover, there is a tendency that regulators are willing to take steps to further regulate or at least a better reporting framework in the FinTech industry[[1]](#footnote-1). Therefore, this study might helpful for regulators to understand statistical properties of joint defaults in P2P lending.

This paper contributes to the credit risk management and FinTech literature in two ways. First, we explore default dependency in P2P lending with a specific focus on asymmetric tail dependency. We conduct an empirical study based on monthly dataset of “Funding Circle” from 08/2010 to 10/2018. The platform operates in several big markets including the UK, US, Germany and the Netherlands. As of 2018, Funding Circle manages more than 5 trillion USD and helps more than 85000 SME reach funding throughout the world[[2]](#footnote-2). The dataset contains detailed information about payment status of six different borrower groups classified on their risky levels. The default rates are calculated by subtracting recovery rates from defaulted amount for a specific lending segment divided by total borrowed amount in that segment.

The second contribution is applying copula methods in the credit risk management of peer to peer lending. Copulas are functions link marginal distributions to the multivariate distributions which have well-defined properties (Nelson, 2006). One of the main advantages of copulas is allowing to identify tail dependence across the multiple distributions. Since the credit risk management literature is particularly interested in assessing the upper (right) tails of the joint default rates distributions, several researches give an emphasis for developing upper tail dependence parameter (Joe, 1997; Trivedi and Zimmer, 2007). In this study, we apply five different copulas (Gaussian, Student’s , Gumbel, Frank, Clayton) to fit joint default distributions. These copulas permit to identify main statistical properties of the loan portfolios. Gumbel copula is able to capture right tail dependence which is a particularly explored aspect of default dependency. Clayton copula is defined to have left tail dependence. Student’s copula captures symmetric tail dependence with equally right and left tail dependence while Gaussian and Frank copulas are defined as symmetric dependence without any tail dependence.

Some clear results emerge from our estimations that are mostly in line with common understandings in credit risk management. First, from the estimations of entire distribution of fifteen portfolio pairs we find that the Gumbel Copula representing right tail dependence fit the empirical data better others in eight out of fifteen portfolios. This result can be interpreted as in eight out of fifteen cases higher default rates are more associated than other quantiles of joint distribution which may inflate losses in the downturn scenarios. In three cases, joint distributions of the loan portfolios exhibit symmetric tail dependence by Student’s t copula meaning that these pairs of loan portfolios are more associated in higher or lower quantiles of distribution rather intermediate rates. In two cases, the loan portfolio pairs are better captured by Clayton copula which exhibits left tail dependence meaning that the pairs are associated with the most profitable scenarios than other quantiles of joint distribution. In other words, borrowers tend to pay back loans and interests simultaneously, while exhibit uncommon behaviour in the downturn scenarios. In two remaining cases, the Frank copula fit portfolio pairs better than others meaning that loan portfolios exhibit symmetric dependence without any tail dependence. Moreover, we find that Gaussian copula fails to fit any of fifteen pairs loan portfolios which is mostly in line with the literature (Crook and Moreira, 2011; Das and Geng, 2015; Di Clemente and Romano, 2004). The loan portfolios can be non-linear and asymmetric and might not be successfully captured by Gaussian copula even if individual distributions are normally distributed.

As the estimation joint extreme default rates is of interest in credit risk management, we carry out the similar estimations for the right tails of joint distributions defined as higher than 75th percentile of the original distributions. We find the similar results as the whole distribution. The copula representing right tailed dependence (Gumbel) are found to be successful in six out of fifteen portfolios pairs meaning that extreme high default dependence is more associated than other quantiles of the joint distributions in the right tails. In three cases, (A+B, A+D and A+ E) the portfolios are better fitted by copulas which means that the tail dependence is symmetrically existent in the higher and lower values defaults than intermediate quantiles. In three cases, (BC and DE) Clayton copulas successfully captures the joint distributions meaning that joint defaults in the upper tailed distribution are more associated in the lower default rates than higher. Finally, the Frank copula fits three of loan portfolio pairs, exposing symmetric tail dependence. As in the previous section, the results demonstrate that Gaussian copula fails to capture the joint defaults in any cases meaning that loan portfolios are heavily right skewed and joint default distributions tend to comove together in the higher values of defaults.

Finally, we compare the normal and copula approaches with the original dataset in predicting joint extreme defaults. Specifically, we calculate the probability of default rates for each of two segments, respectively while simultaneously being above 95th, 90th, 85th, 80th,75th percentiles. In other words, we compare the normal and copula approaches in predicting joint default rates in the highest 5%, 10%, 15%, 20%, 25% of the joint distributions. A better approach is decided based on the difference between estimated default values of dataset and normal vs copula approach. The smaller difference between dataset will be decided as a better method. We find that the copula method outperforms the traditional normal method 52 out of 75 cases (in other words, 70%). The copula method is found to be a better estimator to predict joint extreme defaults and the predicted default rates are found to be closer than those of real dataset. The normal method, however, produce closer estimation in least risky (A+ A) and highest risky (DE) segments. Moreover, the findings do not support the hypothesis that normal method underestimate extreme defaults which are not found as the case in our empirical work. Conversely, we find that in most cases, the normal method estimates considerably higher joint default rates than the original dataset and the copula method. This finding is in line with the findings from credit card portfolios explored in Crook and Moreira (2011).

The remainder of this paper is organized as follows. The second section reviews the literature on default dependency in credit risk management, its models and applications, as well as the key differences between traditional vs P2P lending. In the third section, we present the data sources and the methodologies of the applied copula families. The fourth section discusses the main empirical findings with the whole and right-tailed distribution. We discuss the major findings with concluding remarks in the fifth section.

1. **Literature Review**

**2.1 Default Dependencies**

There has been strong interest in the development of accurate default dependency models. The most obvious reason for concerning about dependency is that it affects the distribution of loan portfolio losses and is therefore critical in determining quantiles or other risk measure used for allocating capital for solvency purposes (Lando, 2004).

Most of the previous literature about modelling default dependency is associated with large portfolios. It can be divided into two main groups: Latent variable and Mixture Models: For example, the models such as KMV (1997), Credit Metrics (1997) have been used for modelling the joint dependency based on relatively simplistic approach (based on multivariate normality). KMV (KMV Corporation, 1997) and Risk Metrics (RiskMatrics Group, 1997) are extensions of the firm value model of Merton (1974) and assumes default happens if obligor’s assets fell below certain level of threshold which is less than liabilities. It assumes certain factor dependency on defaults and is based on Vasicek (1987), Belkin et al., 1998), Finger (1999) and Lucas et al., (2001). In both models, the calibration of the correlation matrix is achieved by using factor model relating changes to the latent variables to systematic changes in a small number of underlying factors (Frey & McNeil, 2003). In a mixture model, default happens under some economic factors such as macroeconomic variables. It assumes independency among default probabilities of obligator’s and dependence is driven by common factors instead (for example, Bernoulli and Poison Mixture models. Bernoulli). Bernoulli Mixture models for two states (default and non-default) can be extended to multinomial mixture model too.

Both latent and Bernoulli mixture models are linked with each other and differs mainly based on interpretation and presentation rather than mathematical ways (Gordy, 2000). These models are used for large portfolio modelling by using factor or latent dependency in defaults. Simplifying assumptions are usually used for individual default dependency. In this case defaults are usually modelled by using copula approach which is based upon a multivariate normal distribution with given marginal.

**2.2 Asymmetric effects**

Traditional credit risk models assume that asset returns are independent from period to period and normally distributed. However, it is observed that asset returns do not follow normal distribution in many cases. Particularly, these weakness challenges the conventional ideas about portfolio diversification and can lead to a quantifiable underestimation of portfolio risk if the returns do not follow any particular distribution.

Non-normality of market returns has been an issue for researchers since 1960s. Moore (1962) documents leptokurtic distribution which was due to existence of too many values near the mean and many out in extreme tails. In addition to these results, Mandelbrot (1963) supports the presence of leptokurtosis and documents that extreme values follow the different mechanism than majority of observations. Fama (1965) also provides an evidence for the non-normality of asset returns. Rosenberg and Schuermann (2006) find the similar trend not just for individual risk, but also for loan portfolios and have confirmed that different risk types follow the different distribution shapes.

Implication of non-linearity to portfolio risk is important to improve efficiency and to prevent from unexpected loss in extreme events. In the last three decades investors have faced with economic downturns (Latin American debt crisis-1980; Recent financial crises 2008-2009) which clear out the fact that such extreme events are happening more frequently than our expectations and their effect to risk management are highly important. Appling traditional models to portfolio risk assumes standard mean, variation and linear relationship between each pair of asset classes. In this case, we assume that relationship between the asset classes are similar to the relationship in the existence of extreme events. This means, we are assuming that joint distribution of asset returns is also normal which indicate the behaviour of portfolio classes together rather than individually. However, in many cases it is observed that expected linear correlation between asset classes follow the different joint behaviour. Hence, assuming linearity under quantify the joint probability during extreme events and over quantify the benefits of portfolio during good periods (JP Morgan, 2009). This is due to asymmetry effect in portfolios joint distribution. For example, Merton (1974) assumes linear correlation where dependency is driven by common factor variable. To overcome this problem Capula theory are used by researchers which allows to model the joint distribution by using marginal distribution (J. P. Nelson, 1999). Modelling default probability by non-linearity assumption rewards the desirable profit as hard as it punishes the extreme movement.

Second form of non-normality relates observing asymmetric dependency which means returns tend to be more dependent in bad times. It has been observed that dependency tend to increase in downturns (left tail events). Ang and Bekaert (2002) and Patton (2006) have confirmed the left tail correlation in stock returns while Ning (2010) have confirmed tail dependency in the equity currency pairs. Similarly, Edwards and Susmel (2001) observed tail dependency in emerging stock market investments and showed a safety-first investor’s behavioural preference is to minimize large losses by considering tail dependency. Left tail events effect the correlation asymmetries on portfolio efficiency, particularly exposure to loss. Due to increased dependency during bad times portfolios may be riskier than normal times. Leibowitz and Bova (2009) find that even well diversified portfolio underperforms during crisis times. Presence of asymmetric dependency in credit asset returns have been observed by Das and Geng (2015) and Di Clemente and Romano (2004). Das and Geng (2004) used a portfolio credit risk model which accounts for asymmetry and tail dependency. Di Clemente and Romano (2004) also take into account non-normality and asymmetric dependency while modelling and optimizing the credit risk of a loan portfolio.

2.3 **Credit risk management in Peer-to-peer versus traditional bank lending**

Peer to peer platforms (P2P) transfers the traditional idea of bank lending to online platforms. However, they do not accept deposits from customers; therefore, take no risk on their balance sheet Deloitte, (2016). P2P also do not accept interest from borrowers or lenders, instead they get fee and commissions from both sides. Investors define their risk appetite and desired maturity for loans and the platforms split the money automatically to several borrowers based on risk appetite of investors.

There are still risks involved however in different dimensions. Investors bear the entire credit risk, as there is no intermediary between borrowers and lenders. On the other side, maturity risk also remains. In the case of early repayments maturity mismatch problem still arises, in which case this maturity risk is also passed to investors. In 2014, $23.7 billion of loans were issued through marketplace platforms globally, concentrated primarily in the US (51%), China (38%) and the UK (10%) (Morgan Stanley, 2015).

In contrast to the orthodox P2P Lending model, in an entirely balance sheet driven model (Balance Sheet Lending) the platform originates the loan (not matching retail or institutional funds to complete funding) and therefore assume the credit risk associated with these loans. They operate with an intermediation model that is more akin to bank lending, by financing loans with equity and debt on their balance sheet and, like banks, periodically refinancing by securitizing pools of the loans they have funded. Retail investors or institutions function as a syndicate, funding the platform’s dedicated balance sheet. Unlike regulated bank lenders, however, these balance sheet model platforms do not have access to deposits to facilitate their lending activity (Zhang et al., 2017).

New online lending platform turns out to be a good deal for borrowers because they get a better interest rate than they might through a traditional bank loan or credit card. However, it is also a good deal for lenders because they earn a higher return than they can through a savings account or certificate of deposit. It reduces the transaction cost by directly matching the borrowers and investor through the Internet. As the overheads, these platforms charge borrowers less and prospect more for the investor. By comparison, banks charge borrowers more to cover the cost such as branch networks. The costs of P2P lending are allegedly only one third of those of commercial banks (The Economist, 2014).

As alternative finance or FinTech lending is a newly developed area and shows high rate of growth rate and penetration rate into different countries, it is highly important to investigate on portfolio credit risk of such platforms. However, several papers have been implied Copula families for traditional finance to model joint default dependencies. For example, Archimedean copula approach is used by Fenech and Shafik (2015) to capture the fat tail distribution and asymmetry behaviour linked to the sensitivity of loss distribution in credit risk management. Similar to the result of Naifar (2011) they have observed that loan portfolio loss cannot be modelled by Gaussian Copula which suffers from asymmetry dependence problem. Hence, Gumbel and Clayton are more suitable and accurate way of modelling default correlation of portfolios. Another paper by Crook and Moreira (2011) estimate the likelihood of joint high default rates in a credit card portfolio by using 10 different ways of copulas (allowing symmetric, right and left tail dependency). It is observed most of the asset pairs are left fail dependent while none of them represented by Gaussian copula.

**2.4** **Copula’s application in credit risk management**

The ways of modelling dependency between asset classes has always been the main focus of portfolio credit risk management. The choice of the better model depends on the dependency structure of obligators and is crucial part of the modelling. One of the ways of modelling dependency is relying on Merton Model (1974) where we assume linear dependency which is not true in most cases. Another way to model dependency is to introduce new modelling mechanism of Copula Theory which helps understand the correlation beyond linearity. McNeil et al., (2005) documented weaknesses of linearity assumption where asset classes follow more complex distribution rather than regular Gaussian multivariate distribution. In contrast, Copula is a function which transfers the multivariate distribution function to its marginal distribution function by quantiles. It is a great tool for modelling dependency where correlation follows the random distribution.

Different models have been used by literature to model dependency. Duffie and Singleton (1999) suggest building default dependency model where joint default events happen at the same time and intensity must be specified for each joint default event which was improved by Kijima and Muromachi (2000). However, the model required to specify the intensities and assuming the default at the same time was unrealistic. Another approach was introduced by Hull and White (2001) to model dependency where default correlations are triggered by firm’s value and correlated with each other. However, the model suffered from implementation problem which was time consuming and the assumption of Gaussian dependency structure of joint distributions was unclear (barriers in the firm’s value effect the joint distribution non-linearly). Poon et al., (2004) have introduced multivariate extreme value theory instead of copulas. However, introduced model is data driven and requires the observed joint tail events.

Copula was first introduced by A. Sklar in 1959 as a statistical mechanism to transfer the joint distribution into its marginals and copula as a model to show the dependency between the marginals. The application of Copula into finance has relatively new history when Embrechts et al., (1999), Li (2000) and Frey and McNeil (2001) applied this technique in their papers. Li (2000) provide implication of copulas to credit risk and credit derivatives and analyse time to default of the different obligators. Frey and McNeil (2001) have confirmed importance of tail dependency by introducing copulas. They applied different default dependence copulas depending on the distribution of returns in loan portfolio. In additional, Copula has been applied to model the dependency of exchange rates Patton (2006), to capture dependency in multivariate contingent claims pricing Rosenberg (2003) and to observe asymmetric dependence in international financial returns (Ang and Bekaert, 2002; Chollete et al., 2009).

There is several Copulas that can be selected. The choice of most appropriate Copula has been an important issue in credit risk management of portfolio which determines the level of dependency. In practice, independent and perfectly correlated variables are naturally inapplicable for copulas. Due to its familiarity, Gaussian Copula was the most famous one among others (for example applied by Aas, 2004; Ward and Lee, 2002), however it failed to capture asymmetry, non-linearity and heavy tail dependency. Therefore, alternative copulas have been used to model joint dependences, for example, Student-t copula which is similar to Gaussian approach, but with positive tail dependency. Student-t copula requires the estimation of ν parameter which is degrees of freedom. As the degrees of freedom increases it approaches to Gaussian copula and useful for investors to observe the effect of fat tails (Bluhm et al., 2003). However, this approach assume symmetry around tails, which is not realistic during extreme events). Hence, Clayton copula is the common proposed one which allow only left tail dependency. In reality, large negative co-movements are more likely to happen than large positive co-movements (University of Oxford, 2016). It is a family of Archimedean Copula, that introduce the generator of ψ which is the function of θ (Equation 1). The parameter of θ controls the dependency in its term. Archimedean family of copulas have been used in literature by Bluhm et al., (2003), Burtschell, et al., (2009) and O’Kane, (2012). Several other ways of Copulas have also been used such as Marshall-Olkin copula, double-t copula and so on. However, we will focus on the most famous Copulas (Gaussian, Student-t and Clayton).

Several papers have been implied Copula families in modelling joint default dependencies in finance. Archimedean copula approach is used by Fenech and Shafik (2015) to capture the fat tail distribution and asymmetry behaviour linked to the sensitivity of loss distribution in credit risk management. Similar to the result of Naifar (2011) they have observed that loan portfolio loss cannot be modelled by Gaussian Copula which suffers from asymmetry dependence problem. Hence, Gumbel and Clayton are more suitable and accurate way of modelling default correlation of portfolios. Another paper by Crook and Moreira (2011) estimate the likelihood of joint high default rates in a credit card portfolio by using 10 different ways of copulas (allowing symmetric, right and left tail dependency). It is observed most of the asset pairs are left fail dependent while none of them represented by Gaussian copula.

To sum up, Copula make a difference to calculate the asset dependency in a portfolio risk management. If the defaults do not follow normal distribution applying Copula will help in calculation of loan loss exposures in a better way. In the case of asymmetric dependence, the choice of right Copula has also been important issue.

1. **Data and Methodology**

**3.1 Data**

We conduct an empirical study based on unexploited monthly data from 08/2010 to 10/2018 of “Funding Circle” which is one the biggest peer-to-peer lending platforms in the UK. It was established in the UK in 2010, now operates in several big markets including the UK, US, Germany and the Netherlands. As of 2018, Funding Circle manages more than 5 trillion USD and helps more than 85000 SME reach funding throughout the world[[3]](#footnote-3). The dataset contains detailed information about payment status of six different borrower groups classified on their risky levels. In the dataset, A+ represents the least risky segment, while E indicates the riskiest segment. Other risk segments are categorized as A, B, C and D.

The credit risk management clearly defines the components of expected loss as probability of default, loss given default and exposure at default Hull (2014). However, the literature is diverse on defining default rate. Moody’s (2005) calculate the default rate for month t as the ratio of all sum of defaulters to number of issuers left in the rating universe. Anson et al., (2004) define default rate par value outstanding for the year to the amount defaulted.

**Figure 1: Pairwise representation of the default rates of each credit segment**







The closest paper to our study Crook and Moreira (2011) calculates default rate as the number of loans that reached their third month in arrears for the first time divided by the total amount of active accounts in that month.

The dataset used in this study covers relatively short time span and availability of enough loans on each segment is an issue for some occasions. Therefore, defining default rate based on number of accounts would introduce distortion and misrepresentation of default rate across different segments. In this study, we follow the methodologies of big lending platforms (including Lending Club, Funding Circle) that involve in indicating ratio of net defaulted payment by total lending mount as default rates. In other words, we find default rates by subtracting recovery rates from defaulted amount for a specific lending segment divided by total borrowed amount in that segment.

Figure 1 show the pairwise representation of the joint default rates of six segments. Table 1 presents the summary statistics of six different portfolios. The mean default rate for B segment is slightly higher than riskier segments such as C, D, E. In the meantime, the default rate of E segment is slightly lower than D. These inconsistencies might be explained by the small sample problem. In the data, we observe that both lending and defaulted amount for E segment is considerably lower than others. Meanwhile, both number and volume of loans are reportedly higher for B segment. Therefore, due to the lack of lending activity in the segment E, default rate is also statistically lower. Similarly, default rate is statistically higher for the segment B due to its higher volume. In the main estimation, we interpolate the data to fill the missing data point to achieve the dataset to be a theoretically representative as much as possible.

The second column in the summary statistics shows the median values. Except segment C, the median and mean values are numerically different from each other that might be the first indication of non-normality in the dataset. Standard deviation values are found to be highest in the segments B and E, which are consistent with the mean default values, the lowest for the segment C. Last two columns demonstrate skewness and kurtosis of all segments. Skewness is found to be close to 0 only in the segment C which is the second indication of non-normality in the dataset except the segment C. All portfolios, except the segment C, expose positive kurtosis suggesting that the segments may have fatter tails than normal distribution.

**Table 1: Summary statistics of default rates of six different segments**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Segment | Mean | Median | Max | Min | Std. dev | Skewness | Kurtosis |
| A+ | 0.0261 | 0.0049 | 0.5922 | 0.0000 | 0.0686 | 6.6581 | 54.3782 |
| A | 0.0553 | 0.0465 | 0.2289 | 0.0000 | 0.0545 | 1.2661 | 1.2962 |
| B | 0.0853 | 0.0604 | 0.7268 | 0.0000 | 0.1031 | 3.4168 | 17.3327 |
| C | 0.0684 | 0.0683 | 0.1767 | 0.0000 | 0.0476 | 0.2330 | -0.7957 |
| D | 0.0760 | 0.0638 | 0.3625 | 0.0000 | 0.0760 | 1.4551 | 2.8403 |
| E | 0.0714 | 0.0283 | 0.4971 | 0.0000 | 0.1071 | 2.4941 | 8.0766 |

To specifically check the normality issue in the data set, we carry out Jarque-Bera test (Jarque and Bera, 1987) which assumes a data is normally distributed under null hypothesis, with an alternative hypothesis of non-normality.

**Table 2: Jarque – Bera test for the default rates’ segments**

|  |  |  |  |
| --- | --- | --- | --- |
| Segment | Jarque – Bera stat. | P value | Null of normality |
| A+ | 10028.45 | 0.001 | Rejected |
| A | 28.08 | 0.015 | Rejected |
| B | 1130.55 | 0.001 | Rejected |
| C | 2.72 | 0.151 | Failed to reject |
| D | 31.47 | 0.000 | Rejected |
| E | 78.29 | 0.000 | Rejected |

Table 2 shows the results of the Jarque-Bera test for six segments. It reports that all portfolio segments, except C, is not normally distributed in the 1% significance level. We fail to reject null of normality in the portfolio C which is consistent with the descriptive statistics in Table 1. Even if individual default rate series are normally distributed, their pairwise joint distribution might not be successfully captured by the Gaussian copula (Crook and Moreira, 2011). Thus, joint distribution of the segment C with other portfolios might not be normally distributed.

* 1. **Methodology**
		1. **Copulas and tail dependency in credit risk**

Copulas are functions link marginal distributions to the multivariate distributions which have well-defined properties (Nelson, 2006). According to Sklar’s theorem (Sklar, 1959), given as a joint distribution function with marginals , there is a copula for all in satisfies the following equation:

 (1)

The theorem indicates that copulas are joint distribution functions, also joint distribution function can be also written as copulas given their marginal distributions. Thus, Schweizer (1991) shows that modelling joint distribution can be reduced to modelling copulas. Since copulas represent the dependence between the variables that result from splitting the joint distribution into a copula and the marginals, copulas are also called dependence functions (Deheuvels, 1978). The probabilistic features of copulas are defined with the following properties. If random and are standard uniformly distributed variables as:

 and (2)

Sklar’s theorem shows that

 (3)

 (4)

 (5)

 (6)

 (7)

 (8)

 (9)

One of the main advantages of copulas is allowing to identify tail dependence across the multiple distributions. Since the credit risk management particularly interested in assessing the upper (right) tails of the default rates distributions, several researches give an emphasis for developing upper tail dependence parameter (Joe, 1997; Trivedi and Zimmer, 2007). The upper tail dependence parameter is given by:

 (10)

Where is the extreme percentile and and are the inverse distribution of and , respectively. In other words, indicates the probability that each of variables is greater than its marginal distribution with the same percentiles. If , the variables satisfy the upper tail dependence condition, whereas there is no upper tail dependence if . The lower tail dependency can be calculated in similar manner but variables smaller than specific cuttoffs when the percentile approaches to zero (Crook and Moreira, 2011). When , the data becomes lower tail dependent.

### **Parameter estimation and model selection techniques**

The simplest approach is to fit copula to data is maximum likelihood which attempts to find the parameter values that maximize the likelihood function given the observations (Brooks, 2008). When the data is transformed to the unit hypercube by parametric estimations of their marginal cumulative distribution functions, it is known as the Inference Functions for Margins method. Alternatively, when the data is transformed by the empirical cumulative distribution function, the method is known as Canonical Maximum Likelihood.

Although, there is no consensus on selecting the best techniques, Genest et al., (2009) emphasizes that Inference Functions for Margins method less efficient as it flaws the estimations of the univariate distributions. While the Canonical Maximum likelihood is a better method for simulated data, Maximum Likelihood method has been extensively used in the statistical applications (Durrleman et al., 2000). Thus, in this study we consider Canonical Maximum Likelihood and Maximum Likelihood methods, we focus on former in the main estimations.

After finding the parameter of each copula, the second step is to select the most optimal copula model that fits empirical data as well as dependence structure as much as possible. As described by (2009) there are a few techniques available to choose the best copula. Some of them, such as information criteria (Akaike and Schwarz’s Bayesian) and Likelihood Ratio test are extensively applied in all areas of statistical and econometrical estimations to decide on the best model choice. Another approach is the one based on distance measures between candidate copula’s and empirical data’s distribution (Kole et al., 2007). This method is also called goodness-of-fit (GoF). Since we do not aim to test different model selection methods, we follow the literature to adopt an appropriate model selection technique.

The simulation experiments show that classes of the GoF method, namely, Empirical copula, Kendall’s transform and Rosenblatt’s transform, present the highest performance (Berg, 2009; Genest et al., 2009). Fitting copulas to empirical data based on three methods would naturally introduce conflicting results. By following Crook and Moreira (2011), we focus on the Empirical Copula method in the final copula selection as it is found as the most reliable estimation technique (Genest et al., 2009) and presents the least data transformation (Berg, 2009).

3.2.3 **Applied copula families**

In this study, we apply five different copula families, namely Gaussian, Student’s t copula and three Archimedean copulas; Frank, Gumbel, Clayton copula. In Table 3, we summarize applied copulas and their main statistical properties. We capture symmetric dependence without tail dependence with Gaussian and Frank copulas, symmetric dependence with upper and lower tail dependence with t copula, left (lower) tail dependence with Clayton copula, right (upper) tail dependence with Gumbel copula.

**Table 3. Applied copulas and their statistical features**

|  |  |
| --- | --- |
| Copula | Dependence Structure |
| Gaussian | Symmetric dependence without tail dependence |
| Frank | Symmetric dependence without tail dependence |
| Student t | Symmetric dependence with (lower and upper) tail dependence |
| Clayton | Left (lower) tail dependence |
| Gumbel | Right (upper) tail dependence |

***Gaussian copula***

Gaussian copula, as the name implies, assumes joint distribution follows bivariate standard normal distribution. Gaussian copula is the most commonly applied copula in finance due to its convenient properties (Meissner, 2014). In the n-variate case, it is defined as:

 (11)

Where is the joint, n-variate cumulative standard normal distribution with , the symmetric, positive-definite correlation matrix of the n-variate normal distribution , is the inverse of a univariate standard normal distribution.

In the meantime, Cherubini et al., (2004) show that if the are uniform, then the are standard normal and is standard multivariate normal distribution. Equation (11) is a general description of Gaussian copula. David Lee(2000) default probabilities is defined for entity , at a fixed time , as marginal distributions. Thus, the following Gaussian default time copula ,

 (12)

According to Cherubini et al., (2004), equation (12) can be interpreted as: Given the respective marginal distributions in which the cumulative default probabilities of entities at times , A Gaussian copula function should accordingly exist that enables mapping of the marginal distributions via to standard normal and the joining of the to a single n-variate standard normal distribution with the correlation structure .

In other words, the term in equation (12) maps the cumulative default probabilities of asset , for time , percentile to percentile a univariate standard normal distribution. For example, 5th percentile of is mapped to the 5th percentile of the standard normal distribution, the 10th percentile of is mapped to the 10th percentile of the standard normal distribution.

 **copula**

Similar to Gaussian copula, Student’s copula is also a symmetric, but can capture upper and lower tail dependence as it follows the properties of distribution. copula maps marginal distributions to the distribution by percentile to percentile base. In recent years, copula has received a significant attention in modelling multivariate financial returns data such as finding joint distribution of stock returns (Embrechts et al., 1999; Fang, Fang and Kotz, 2002). Empirical studies show that copula performs significantly better Gaussian copula due to its favourable statistical properties such as the ability to better capture dependent extreme values usually observed in finance (Breymann et al.,, 2003; Mashal and Zeevi, 2002).

The Student’s copula is defined as follows:

 (13)

Where is the covariance matrix, is the standardized multivariate Student’s distribution with covariance matrix and degree of freedom. indicates the inverse of Student’s cumulative distribution function. The main advantage of Student’s copula over Gaussian copula is assuming a non-zero tail dependence even if correlation is zero. In the bivariate case of the Student’s copula, the formula can be defined as:

 (14)

Where is the correlation coefficient. Tail dependency increases with an increase in the degree of freedom. As long as the condition satisfies, Student’s copula approaches to the Gaussian copula.

**Archimedean copulas**

According to Cherubini et al., (2004) Archimedean copulas can be constructed by specifying a particular generator function , such that satisfies the following conditions. Given a strict generator , and is completely monotonic on , then bivariate Archimedean copula can be defined as:

 (15)

Accordingly, the dependence measure via Kendal’s tau can be given:

 (16)

Cherubini et al., (2004) demonstrates that the coefficients of upper tail dependence and lower tail dependence can be given by:

 (17)

And

 (18)

***Clayton copula***

The Clayton copula, mostly used to estimate lower tail dependency is introduced by Clayton (1978). The closed form expression of the bivariate Clayton copula is given by:

 (19)

where is the copula parameter restricted on the interval . If , then the marginal distributions become independent; when , then the Clayton copula approaches to the Frechet - Hoeffding upper bound. The relationship between the Clayton copula parameter and Kendall’s tau rank correlation is given by:

 (20)

According to Cherubini et al., (2004), the parameter of lower tail dependency for this copula can be calculated by:

 (21)

***Gumbel copula***

The Gumbel copula is developed by (Gumbel, 1960) to capture strong upper tail dependence and weak lower tail dependence. If the bivariate distribution is expected to be jointly dependent at high values and less dependent at low values, then the Gumbel copula might be an appropriate choice. The bivariate Gumbel copula is defined by:

 (22)

Where is the copula parameter restricted on the interval . When approaches 1, the marginals become independent and when goes to infinity, the Gumbel copula approaches the Frechet-Hoeffding upper bound. As the Clayton copula, the Gumbel copula represents only independent and positive dependence.

The relationship between the Gumbel copula parameter and the Kendall’s tau is given by:

 (23)

Accordingly, the parameter of the upper and lower tail dependence of the Gumbel copula can be estimated by:

 (24)

***Frank copula***

Unlike the Clayton and Gumbel copula, Frank copula (Frank, 1979) allows both positive and negative dependence in data. The general formula can be expressed by:

 (25)

Where is the copula parameter that can take any value. Meantime, approaches and , the Frank copula will approximate Frechet-Hoeffding in both upper and lower bound. Conversely, when approaches 0, the formula will be in the independent case. However, unlike the Clayton and the Gumbel copulas, the Frank copula does not capture upper or lower tail dependence (). Therefore, the Frank is the most suitable in modelling data with weak tail dependence. The calculation of Kendal’s tau, first requires the calculation of Debye function which is given by:

 (26)

Afterwards, tau is given by:

 (27)

1. **Empirical results**

**4.1 Fitting copula to the entire distribution**

To estimate dependency structure between credit portfolios, we apply the underlying copula methods to the whole dataset. Since we have six risky segments, copulas need to be fit on twelve pairs of portfolios. As specified in the previous sections, we primarily focus on the Canonical Maximum Likelihood method to estimate the copula parameters as it is shown to outperform other methods (see section 3). Then we use the Empirical Copula Method as a

goodness-of-fit measure to determine which copula method better represents the dependence structure.

Table 4 reports the estimated parameters for each copula across all pairs of loan portfolios. Table 5 presents the dependence measures of fifteen pairs of risky loan portfolios (in lower triangle) and the copulas are found to be best fit to the data (in upper triangle). The table reveals that eight out of fifteen risky segments denote right tail dependence (i.e ) meaning that higher default rates are more associated than other quantiles of joint distributions which may inflate losses in the downturn scenarios. In this study, right tail dependency is captured by Gumbel copula.

In three cases, (pairs, AB, A+D and CE) joint distributions of the loan portfolios exhibit symmetric tail dependence (i.e ) by Student’s t copula. It indicates that these pairs of loan portfolios are more associated in higher or lower quantiles of distribution rather intermediate rates. Two of the loan portfolio pairs (BC and A+E) are better captured by Clayton copula which exhibits left tail dependence () meaning that the pairs are associated with the most profitable scenarios than other quantiles of joint distribution. In other words, borrowers tend to pay back loans and interests simultaneously, while exhibit uncommon behaviour in the downturn scenarios. Finally, in two remaining cases, Frank copula fits portfolio pairs meaning that the pairs exhibit symmetric dependence without tail dependence.

Consistent with the previous finding (Crook and Moreira, 2011), even if individual distributions (in our case, portfolio C) are normally distributed, their joint distributions might not be successfully captured by Gaussian copula. In fact, Table 5 shows that Gaussian copula fails to fit any of fifteen pairs loan portfolios. Our baseline results are in line with the findings of the traditional credit risk management literature. Specifically, our evidence for non-

**Table 4. Copula parameters for fifteen pairs of loan portfolios based on the entire distribution**

|  |  |  |
| --- | --- | --- |
| GoF measure | Copula | Parameters |
| A+ A | A+B | A+ C | A+ D | A+ E | A B | A C |  A D |
| Empirical copula | Gaussian | 0.26 | 0.04 | 0.24 | 0.11 | 0.12 | 0.54 | 0.19 | 0.4 |
| Student t | 0.95 | 0.91 | 0.95 | 0.86 | 0.87 | 0.96 | 0.96 | 0.92 |
| Gumbel | 7.26 | 4.72 | 6.58 | 3.35 | 3.44 | 7.51 | 8.14 | 4.97 |
| Frank | 23.72 | 13.23 | 21.57 | 9.05 | 9.68 | 23.45 | 26.85 | 14.49 |
| Clayton | 4.09 | 2.99 | 3.63 | 1.98 | 1.97 | 5.01 | 4.58 | 3.07 |

|  |  |  |
| --- | --- | --- |
| GoF measure | Copula | Parameters |
| A E |  B C | B D | B E | C D | C E | D E |
| Empirical copula | Gaussian | 0.34 | -0.04 | 0.46 | 0.33 | 0.12 | 0.34 | 0.53 |
| Student t | 0.92 | 0.94 | 0.93 | 0.93 | 0.91 | 0.93 | 0.97 |
| Gumbel | 4.99 | 5.62 | 5.33 | 5.21 | 4.37 | 5.18 | 7.91 |
| Frank | 15.09 | 17.06 | 15.42 | 15.45 | 12.59 | 16 | 23.78 |
| Clayton | 2.99 | 3.58 | 3.41 | 3.54 | 2.68 | 3.08 | 5.38 |

**Table 5. Best-fit copulas based on the entire distribution**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | A+ | A | B | C | D | E |
| A+ | 1 | Gumbel | Gumbel | Gumbel | Student’s t | Clayton |
| A | 0.18080.2698 | 1 | Student’s t | Gumbel | Gumbel | Frank |
| B | 0.12670.004 | 0.13040.006 | 1 | Clayton | Frank | Gumbel |
| C | 0.25350.2415 | 0.23920.1316 | 0.0958-0.0783 | 1 | Gumbel | Student’s t |
| D | -0.0840.1018 | 0.28760.377 | 0.4210.3926 | 0.05050.0923 | 1 | Gumbel |
| E | 0.06270.0852 | 0.28610.2701 | 0.37420.237 | 0.22830.3002 | 0.48560.4336 | 1 |

*Notes: Best-fit copulas (upper-right triangle) and dependence measures (lower-left triangle) for joint default rates based on the entire distributions. The dependence measures are the linear correlation (above) and the Kendall’s tau in lower. The fitted copulas are selected with Empirical copula based on the Canonical Maximum Likelihood functions*

normality and tail dependency are supported by Das and Geng (2015), Di Clemente and Romano (2004).

**4.2 Fitting copulas to the right tails**

Since the estimation of joint high default rates is of interest in the credit risk management, we carry out the previous estimation based on the upper tails of the default rates distributions. Specifically, we estimate the copulas based on the right tails (defined as above the 75th percentile of the original default distribution). In this manner, the selected distribution of each risky segment represents 25% highest default points. Afterwards, each candidate copula is defined as:

where is a candidate copula with a parameter estimated with the same procedures as in the previous section but limited to percentiles above 0.75.

Table 6 presents the estimated parameters for the candidate copulas based on fifteen different portfolios. Table 7 presents the results of the estimations that choose the best copulas fit the joint default distribution in the right tails. The results are mostly in line with the estimations carried out with the whole distribution. The copula representing right tailed dependence (Gumbel) is found to be successful in six out of fifteen portfolios pairs. It means that extreme high default dependence is more associated than other quantiles of the joint distribution. In three cases, (A+B, A+D and A+ E) the portfolios are better fitted by copulas which means that the tail dependence is symmetrically existent in the higher and lower values defaults than intermediate quantiles. In three cases, (BC and DE) Clayton copulas successfully captures the joint distribution meaning that joint defaults in the upper tailed distribution are more associated in the lower default rates than higher. Finally, in three remaining cases, Frank copula fits portfolio pairs meaning that the pairs exhibit symmetric dependence without tail dependence. As in the previous section, the results demonstrate that Gaussian copula fails to capture the joint defaults in any cases meaning that loan portfolios are heavily right skewed and joint default distributions tend to comove together in the higher values of defaults.

**Table 6. Copula parameters for fifteen pairs of loan portfolios based on the upper tail of the default rate distribution**

|  |  |  |
| --- | --- | --- |
| GoF measure | Copula | Parameters |
| A+ A | A+B | A+ C | A+ D | A+ E | A B | A C |  A D |
| Empirical copula | Gaussian | 0.0371 | 0.7987 | 0.0829 | 0.0902 | 0.0987 | 0.1288 | -0.138 | 0.1742 |
| Student t | 0.922 | 0.8895 | 0.9358 | 0.8479 | 0.8755 | 0.9124 | 0.9844 | 0.9578 |
| Gumbel | 4.6595 | 3.5358 | 5.2537 | 2.8439 | 3.3608 | 4.4119 | 10.7442 | 6.0122 |
| Frank | 13.5013 | 9.9895 | 16.0216 | 7.5353 | 9.1005 | 2.9465 | 32.1348 | 17.6316 |
| Clayton | 3.2465 | 2.4237 | 3.5708 | 1.8457 | 2.3602 | 4.1899 | 7.3917 | 4.3409 |

|  |  |  |
| --- | --- | --- |
| GoF measure | Copula | Parameters |
| A E |  B C | B D | B E | C D | C E | D E |
| Empirical copula | Gaussian | 0.2038 | 0.0253 | 0.0903 | 0.0352 | 0.133 | 0.1956 | 0.9717 |
| Student t | 0.9746 | 0.9257 | 0.9346 | 0.9446 | 0.9594 | 0.9724 | 0.9892 |
| Gumbel | 8.2998 | 4.3802 | 4.9381 | 4.5241 | 5.967 | 8.4443 | 11.2489 |
| Frank | 24.7465 | 12.734 | 14.8506 | 13.3462 | 17.2835 | 24.9277 | 32.201 |
| Clayton | 6.284 | 4.3084 | 4.938 | 4.9724 | 4.432 | 6.6408 | 9.3161 |

**Table 7: Best fit copulas based on the upper tail of the default rate distribution**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | A+ | A | B | C | D | E |
| A+ | 1 | Gumbel | Student’s t | Gumbel | Student’s t | Student’s |
| A | 0.15150.0054 | 1 | Gumbel | Clayton | Gumbel | Gumbel |
| B | 0.09090.8137 | 0.02160.1356 | 1 | Clayton | Frank | Frank |
| C | 0.06490.0503 | -0.0563-0.1501 | 0.08230.0171 | 1 | Gumbel | Frank |
| D | 0.30580.0489 | 0.33980.1644 | 0.11890.081 | 0.0680.1235 | 1 | Clayton |
| E | 0.40230.1805 | 0.33170.1987 | 0.000.285 | 0.10430.205 | 0.87710.9741 | 1 |

*Notes: Best-fit copulas (upper-right triangle) and dependence measures (lower-left triangle) for joint default rates based on the entire distributions. The dependence measures are the linear correlation (above) and the Kendall’s tau in lower. The fitted copulas are selected with Empirical copula based on the Canonical Maximum Likelihood functions*

**4.3 Estimation of joint extreme default rates**

In the final estimations, we compare the normal and copula approaches with the original dataset in predicting joint extreme defaults. Specifically, we calculate the probability of default rates, and in segments and , respectively, while being simultaneously above specific levels (values) and as follows:

* Assuming normality, joint extreme default rates can be calculated as: where indicates the cumulative distribution of a normal distribution
* By using copula approach, joint extreme default rates can be calculated: where is a survival copula ranks survival rates as ; and are the cumulative distribution function of the original default rate distribution, and , respectively.

We define extreme joint default rates in five categories, 95th, 90th, 85th, 80th,75th percentiles. In other words, we compare the normal and copula approaches in predicting joint default rates in the highest 5%, 10%, 15%, 20%, 25% of the distributions. For instance, we calculate the probability that the 5% highest default rates in segment happen simultaneously the 5% highest default rates in segment .

Table 8 demonstrate the results of estimating extreme joint default rates by the dataset, normal and copula methods in five different percentiles. The first column shows the percentiles of the estimations. For example, in 95th, we estimate the probability that the 5% of highest default rates of both segments happen at the same time. The second column indicates that the proportion of the number of simultaneous default rates for a specific percentile in the whole observations. The third and fourth columns exhibit extreme joint default rates for a specific percentile estimated by normal and copula methods, respectively. The superior method whether normal or copula estimate extreme joint defaults is decided based on absolute differences between the estimated joint default values of normal vs dataset and copula vs dataset. The smaller the difference between dataset will be decided as a better method.

The results indicate that the copula method outperforms the traditional normal method 52 out of 75 cases (in other words, 70%). The copula method is found to be a better estimator to predict joint extreme defaults and the predicted default rates are found to be closer than those of real dataset. The normal method, however, produce closer estimation in least risky (A+ A) and highest risky (DE) segments. Moreover, the findings do not support the hypothesis that normal method underestimate extreme defaults which are not found as the case in our empirical work. Conversely, we find that in most cases, the normal method estimates considerably higher joint default rates than the original dataset and the copula method. This finding is in line with the findings from credit card portfolios explored in Crook and Moreira (2011).

**Table 8. Comparison of predicted default rates based on the entire distribution**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Quantiles | A+ A (Gumbel) | A+ B (Gumbel) | A+ C (Gumbel) | A+ D (Student’s t) | A+ E (Clayton) |
| Dataset | Normal | Copula | Dataset | Normal | Copula | Dataset | Normal | Copula | Dataset | Normal | Copula | Dataset | Normal | Copula |
| 95 | 0.01149 | 0.13911 | 0.14436 | 0 | 0.18215 | 0.04200 | 0.0114 | 0.13814 | 0.09318 | 0.02298 | 0.14956 | 0.13365 | 0.0119 | 0.16983 | 0.08114 |
| 90 | 0.01149 | 0.14916 | 0.16292 | 0 | 0.18864 | 0.05388 | 0.03458 | 0.16878 | 0.11184 | 0.06897 | 0.15221 | 0.15901 | 0.0574 | 0.16983 | 0.09736 |
| 85 | 0.03458 | 0.15313 | 0.16886 | 0.03458 | 0.19208 | 0.06229 | 0.04598 | 0.17248 | 0.12310 | 0.06897 | 0.15296 | 0.18006 | 0.1149 | 0.17417 | 0.11753 |
| 80 | 0.05747 | 0.1543 | 0.17695 | 0.06897 | 0.19209 | 0.08116 | 0.06897 | 0.17248 | 0.12635 | 0.14943 | 0.15715 | 0.19063 | 0.1609 | 0.18476 | 0.12471 |
| 75 | 0.10345 | 0.1543 | 0.18898 | 0.08046 | 0.19209 | 0.08702 | 0.10345 | 0.17248 | 0.13498 | 0.06896 | 0.1588 | 0.19609 | 0.1724 | 0.18532 | 0.13181 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Quantiles | A B (Student’s t) | A C (Gumbel) | A D (Gumbel) | A E (Frank) | B C (Clayton) |
| Dataset | Normal | Copula | Dataset | Normal | Copula | Dataset | Normal | Copula | Dataset | Normal | Copula | Dataset | Normal | Copula |
| 95 | 0.02298 | 0.01954 | 0.06833 | 0.0114 | 0.01779 | 0.05222 | 0.02298 | 0.0117 | 0.01691 | 0.0119 | 0.0138 | 0.01184 | 0.0114 | 0.17636 | 0.05413 |
| 90 | 0.05747 | 0.03451 | 0.10534 | 0.01149 | 0.06128 | 0.05942 | 0.06897 | 0.01621 | 0.05763 | 0.05748 | 0.01814 | 0.04272 | 0.0114 | 0.20884 | 0.06010 |
| 85 | 0.06889 | 0.04168 | 0.11471 | 0.02299 | 0.0773 | 0.08233 | 0.11494 | 0.02225 | 0.07971 | 0.11494 | 0.03001 | 0.05751 | 0.0229 | 0.23025 | 0.07601 |
| 80 | 0.09195 | 0.04495 | 0.13881 | 0.06897 | 0.08426 | 0.08736 | 0.06897 | 0.02314 | 0.09683 | 0.14943 | 0.03259 | 0.06803 | 0.0459 | 0.24191 | 0.07959 |
| 75 | 0.14943 | 0.04629 | 0.14714 | 0.09195 | 0.08445 | 0.09209 | 0.14943 | 0.02314 | 0.10566 | 0.19540 | 0.03693 | 0.07408 | 0.0459 | 0.24819 | 0.08939 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Quantiles | B D (Frank) | B E (Gumbel) | C D (Gumbel) | C E (Student’s t) | D E (Gumbel) |
| Dataset | Normal | Copula | Dataset | Normal | Copula | Dataset | Normal | Copula | Dataset | Normal | Copula | Dataset | Normal | Copula |
| 95 | 0.02298 | 0.17696 | 0.03147 | 0.0119 | 0.18474 | 0.01774 | 0.02298 | 0.01113 | 0.01791 | 0.0119 | 0.01384 | 0.00938 | 0.0119 | 0.05911 | 0.02383 |
| 90 | 0.06897 | 0.18256 | 0.06010 | 0.05748 | 0.18975 | 0.03950 | 0.03448 | 0.01564 | 0.03074 | 0.05747 | 0.02525 | 0.01787 | 0.0689 | 0.06467 | 0.02383 |
| 85 | 0.09195 | 0.18982 | 0.07602 | 0.11494 | 0.2122 | 0.05359 | 0.02298 | 0.02504 | 0.05109 | 0.11494 | 0.02798 | 0.02398 | 0.1194 | 0.06467 | 0.02383 |
| 80 | 0.07478 | 0.19447 | 0.07959 | 0.16092 | 0.21703 | 0.05924 | 0.05747 | 0.03687 | 0.05302 | 0.17241 | 0.03091 | 0.04053 | 0.1609 | 0.06467 | 0.02383 |
| 75 | 0.17241 | 0.19872 | 0.08994 | 0.20689 | 0.21843 | 0.06905 | 0.06897 | 0.03835 | 0.05663 | 0.19540 | 0.03203 | 0.04626 | 0.2069 | 0.06467 | 0.02383 |

*Notes: Comparison of normal and copula method in predicting extreme joint default rates. The copulas are estimated based on the best fitting copulas from Table 5*

1. **Concluding Remarks**

We explore default dependency in P2P lending. According to recent survey, default risk is one of the major concern among the UK platforms (Zhang et al., 2018). It opens a new challenge for researchers to explore the empirical evidence from P2P lending. While P2P lending is considered less risky than traditional banking, its novel nature and lack of experience in the stress scenarios might be indications to be less confident regarding the risk framework of P2P lending. Strong systematic dependency between obligators can create a huge systematic risk even if individual defaults are not particularly concerning. Although default dependency and its asymmetric properties have been widely explored in the traditional lending (Servigny, 2004; Frey, McNeil and Nyfeler, 2001; Frey and McNeil, 2002; Frey, McNeil and Nyfeler, 2003), the similar studies have not been carried out for the P2P lending as it is a relatively new phenomenon in financial markets. To capture possible non-linearity and asymmetric features, we apply five different copula families to empirically fit loan portfolios, as well as, to compare with the performance of the traditional normal method in predicting joint extreme defaults.

We conduct an empirical study based on monthly dataset of “Funding Circle” from 08/2010 to 10/2018. The platform manages more than 5 trillion USD and helps more than 85000 SME reach funding by operating in several big markets including the UK, US, Germany and the Netherlands. The dataset contains detailed information about payment status of six different borrower groups classified on their risky levels. We estimate five different copula families for fifteen portfolio pairs in the entire and upper tailed distributions. This study contributes to the credit risk management and FinTech literature with the following findings.

From the estimations of the entire distributions of fifteen portfolio pairs we find that the Gumbel Copula representing right tail dependence fit the empirical data better others in eight out of fifteen portfolios meaning that higher default rates are more associated than other quantiles of joint distribution. In three cases, joint distributions of the loan portfolios exhibit symmetric tail dependence by Student’s t copula meaning that these pairs of loan portfolios are more associated in higher or lower quantiles of distribution rather intermediate rates. In two cases, the loan portfolio pairs are better captured by Clayton copula which exhibits left tail dependence meaning that the pairs are associated with the most profitable scenarios than other quantiles of joint distribution. In two remaining cases, the Frank copula fit portfolio pairs better than others meaning that loan portfolios exhibit symmetric dependence without any tail dependence.

Next, we carry out the similar estimations for the right tails of joint distributions defined as higher than 75th percentile of the original distributions. The Gumbel copula representing right tailed dependence is found to be successful in six out of fifteen portfolios pairs. In three cases, the portfolios are better fitted by copulas. In three cases, Clayton copulas successfully captures the joint distributions. Finally, the Frank copula fits three of loan portfolio pairs, exposing symmetric tail dependence.

Finally, we compare the normal and copula approaches with the original dataset in predicting joint extreme defaults. Specifically, we calculate the probability of default rates for each of two segments, respectively while simultaneously being above 95th, 90th, 85th, 80th,75th percentiles. We find that the copula method outperforms the traditional normal method 52 out of 75 cases (in other words, 70%). The normal method, however, produce closer estimation in least risky (A+ A) and highest risky (DE) segments. Moreover, the findings do not support the hypothesis that normal method underestimate extreme defaults. Conversely, we find that in most cases, the normal method estimates considerably higher joint default rates than the original dataset and the copula method.

This study suggests to the credit risk management literature that it is not only the default risk concerning factor in P2P lending. Default dependency, in particular, right tailed dependency in P2P lending is found to be a worthwhile area to pay attention as much as traditional banking. The main shortcoming feature of this study is carrying out empirical analysis with relatively fewer data point (87 months) due to the fact Funding Circle has been operating only since 2010. Therefore, a natural extension of this study would be covering a longer time-series data after some time lapsing or applying multiple platforms simultaneously. Another approach to the data insufficiency would be simulating the original dataset to have a better representative dataset. This study may also be interesting for regulators as they are willing to take further steps to regulate or at least to establish a better reporting mechanism for the FinTech industry. As this industry has been tested via a crisis, this study might be helpful to understand that P2P lending carries almost the similar risks as traditional banking even if they are supposedly less risky.

**Appendix**

**Table 9: Logistic regression results**

**Note:** We conduct an empirical study based on unexploited monthly data from 08/2010 to 10/2018 of “Funding Circle” which is one the biggest peer-to-peer lending platforms in the UK. Credit Band E is the riskiest class of loans and A+ is loan class with the lowest risk of borrowers. Dependent variable is the default variable which is equal to 1 if the borrower default and 0 in all the other cases (late, repaid or prepaying stage). We do not report insignificant variables which were different regions, business type that the lenders reported when they applied for business loans, and repayment type which changes between amortizing and interest only. Interestingly loan amount was also indicated insignificant result and not included in the table. To control the time changes, years are added as dummy variable.

\*\*\*represent significance level at the 1% level \*\* indicates significance at 5%level.

|  |  |  |
| --- | --- | --- |
| **Variable names** | **Coefficients** | **Std. Err.** |
| 12 months | 14.6541\*\*\*[0.000] | 0.4706 |
| 24 months | 15.4644\*\*\*[0.000] | 0.4499 |
| 36 months | 13.6338\*\*\*[0.000] | 0.3908 |
| 48 months | 6.9965\*\*\*[0.000] | 0.2306 |
| Payments remaining | 0.58844\*\*\*[0.000] | 0.0154 |
| Number of loan parts | 0.0002\*\*[0.045] | 0.0001 |
| A (low risk) | 0.8816\*\*\*[0.000] | 0.1431 |
| B (below average) | 1.1067\*\*\*[0.000] | 0.1471 |
| C (average) | 1.3937\*\*\*[0.000] | 0.1523 |
| D | 1.5423\*\*\*[0.000] | 0.1700 |
| E | 1.5624\*\*\*[0.000] | 0.2378 |
| Limited Company | 0.3757[0.000] | 0.2082 |
| Limited Liability Partnership (LLP) | 0.4104[0.000] | 0.5254 |
| Partnership partners/LLPs≥4 | 0.3807[0.000] | 0.7812 |
| Partnership>4 partners) | -1.0818\*\*[0.017] | 0.4516 |
| Whole/Partial loan | -0.3279\*\*[0.005] | 0.1165 |
| Repayment type (Amortizing/Interest only) | 0.9774[0.000] | 0.6441 |
| Constant | -18.1437\*\*\*[0.000] | 1.0220 |
| Number of observations | 39,580 |
| Pseudo  | 0.6477 |

Business Type

comparable to private company

Credit Bands

comparable to A+

Maturity

comparable to 60 months

# **Bibliography**

Aas, K. (2004). *Modelling the dependence structure of financial assets: A survey of four copulas*.

Ang, A., & Bekaert, G. (2002). International Asset Allocation With Regime Shifts. *Review of Financial Studies*, *15*(4), 1137–1187.

Anson, M., Fabozzi, F., Choudhry, M., & Chen, R.-R. (2004). *Credit derivatives : instruments, applications and pricing*. Wiley.

Belkin, B., Suchower Stephan, & Forest, R. L. (1998). *The effeft of systematic credit risk on loan portfolio value-at-risk and loan pricing*.

Berg, D. (2009). Copula goodness-of-fit testing: an overview and power comparison. *The European Journal of Finance*, *15*(7–8), 675–701.

Bluhm, C., Overbeck, L., Wagner, C., & CRC Press. (2003). *An introduction to credit risk modeling*. Chapman & Hall/CRC.

Breymann, W., Dias, A., & Embrechts, P. (2003). Dependence Structures for Multivariate High-Frequency Data in Finance. *Quantitative Finance* , *3*(1–14). Retrieved from www.math.ethz.ch/finance

Brooks, C. (2008). *Introductory econometrics for finance*. Cambridge University Press.

Burtschell, X., Gregory, J., & Laurent, J.-P. (2009). A Comparative Analysis of CDO Pricing Models under the Factor Copula Framework. *The Journal of Derivatives*, *16*(4), 9–37.

Cherubini, U., Luciano, E., & Vecchiato, W. (2004). *Copula Methods in Finance*. Oxford, UK: John Wiley and Sons Ltd.

Chollete, L., Heinen, A., & Valdesogo, A. (2009). Modeling International Financial Returns with a Multivariate Regime-switching Copula. *Journal of Financial Econometrics*, *7*(4), 437–480.

Clayton, D. G. (1978). A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence. *Biometrika*, *65*(1), 141.

Crook, J., & Moreira, F. (2011). Checking for asymmetric default dependence in a credit card portfolio: A copula approach. *Journal of Empirical Finance*, *18*(4), 728–742.

Das, S. R., & Geng, G. (2004). Correlated Default Processes: A Criterion-Based Copula Approach. *Journal of Investment Management*, *2*(2).

Das, S. R., & Geng, G. (2015). Correlated Default Processes: A Criterion-Based Copula Approach. In *The Credit Market Handbook* (pp. 186–218). Hoboken, NJ, USA: John Wiley & Sons, Inc.

Deheuvels, P. (1978). *Caracterisation complete des Lois Extremes Multivariees et de la Convergence des Types Extremes*. Retrieved from https://tel.archives-ouvertes.fr/file/index/docid/656468/filename/these.pdf

Deloitte. (2016). *A temporary phenomenon? Marketplace lending*. Retrieved from https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/financial-services/deloitte-uk-fs-marketplace-lending.pdf

Di Clemente, A., & Romano, C. (2004). *Measuring and Optimizing Portfolio Credit Risk: A Copula-based Approach\**. *Economic Notes* (Vol. 33). Wiley/Blackwell (10.1111).

Duffie, D., & Singleton, K. (1999). Modeling Term Structures of Defaultable Bonds. *Review of Financial Studies*, *12*(4), 687–720.

Durrleman, V., Nikeghbali, A., & Roncalli, T. (2000). Which Copula is the Right One? *SSRN Electronic Journal*.

Edwards, S., & Susmel, R. (2001). Volatility dependence and contagion in emerging equity markets. *Journal of Development Economics*, *66*(2), 505–532.

Embrechts, P., Mcneil, A., & Straumann, D. (1999). Correlation and dependence in risk management: Properties and pitfalls. *Risk Management: Value at Risk and Beyond*, 176--223.

Fama, E. F. (1965). The Behavior of Stock-Market Prices. *The Journal of Business*, *38*(1), 34–105.

Fang, H.-B., Fang, K.-T., & Kotz, S. (2002). The Meta-elliptical Distributions with Given Marginals. *Journal of Multivariate Analysis*, *82*(1), 1–16.

Fenech, J. P., Vosgha, H., & Shafik, S. (2015). Loan default correlation using an Archimedean copula approach: A case for recalibration. *Economic Modelling*, *47*, 340–354.

Finger, C. (1999). *Conditional approaches for CreditMetrics portfolio distributions*.

Frank, M. J. (1979). On the Simultaneous Associativity of F (x, y) and x+y-F(x,y). *Equtiones Mathematice*. Retrieved from https://eudml.org/doc/136825

Frey, R., & McNeil, A. (2003). Dependent defaults in models of portfolio credit risk. *The Journal of Risk*, *6*(1), 59–92.

Frey, R., & McNeil, A. J. (2001). Modelling dependent defaults.

Genest, C., Rémillard, B., & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, *44*(2), 199–213.

Gordy, M. B. (2000). A comparative anatomy of credit risk models. *Journal of Banking & Finance*, *24*(1–2), 119–149.

Gumbel, E. J. (1960). Copula Distribution des valeurs extremes en plusieurs dimensions in finance. *Publications de I’Institut de Statistique de I’Universite de Paris*, *9*, 171–173.

Hull, J. (2014). *Risk management and financial institutions*.

Hull, J., & White, A. (2001). The General Hull–White Model and Supercalibration. *Financial Analysts Journal*, *57*(6), 34–43.

J. P. Morgan. (2009). Non-normality of Market Returns. *J.P.Morgan*.

Jarque, C. M., & Bera, A. K. (1987). A Test for Normality of Observations and Regression Residuals. *International Statistical Review / Revue Internationale de Statistique*, *55*(2), 163.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*.

Kijima, M., & Muromachi, Y. (2000). Credit Events and the Valuation of Credit Derivatives of Basket Type. *Review of Derivatives Research*, *4*(1), 55–79.

KMV Corporation. (1997). *KMV Credit Risk Modeling*.

Kole, E., Koedijk, K., & Verbeek, M. (2007). Selecting copulas for risk management. *Journal of Banking & Finance*, *31*(8), 2405–2423.

Lando, D. (2004). *Credit risk modeling : theory and applications*. Princeton University Press.

Leibowitz, M. L., & Bova, A. (2009). Diversification Performance and Stress-Betas. *The Journal of Portfolio Management*, *35*(3), 41–47.

Li, D. X. (2000). On Default Correlation: A Copula Function Approach. *The Journal of Fixed Income*, *9*(4), 43–54.

Lucas, A., Klaassen, P., Spreij, P., & Straetmans, S. (2001). An analytic approach to credit risk of large corporate bond and loan portfolios. *Journal of Banking & Finance*, *25*(9), 1635–1664.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *The Journal of Business*, *36*(4), 394.

Mashal, R., & Zeevi, A. (2002). Beyond Correlation: Extreme Co-movements Between Financial Assets. *SSRN Electronic Journal*.

McNeil, A. J., Frey, R., & Embrechts, P. (2005). *Quantitative risk management : concepts, techniques and tools*.

Meissner, G. (2014). *Correlation risk modeling and management : an applied guide including the Basel III correlation framework-- with interactive models in Excel/VBA*. Wiley.

Merton, R. C. (1974). On the pricing of Corporate Debt: The Risk Structure of Interest Rates. *The Journal of Finance*, *29*(2), 449–470.

Moody’s. (2005). *Defaults, Losses and Rating Transitions*.

Moore, P. G. (1962). Regression As An Analytical Tool. *Applied Statistics*, *11*(2), 106.

Morgan Stanley. (2015). Can P2P Lending Reinvent Banking? | Morgan Stanley. Retrieved November 2, 2017, from http://www.morganstanley.com/ideas/p2p-marketplace-lending

Naifar, N. (2011). Modelling dependence structure with Archimedean copulas and applications to the iTraxx CDS index. *Journal of Computational and Applied Mathematics*, *235*(8), 2459–2466.

Nelson, J. P. (1999). Consumer bankruptcy and chapter choice: State panel evidence. *Contemporary Economic Policy*, *17*(4), 552–566.

Nelson, R. B. (2006). *An Introduction to Copulas*. Springer New York.

O’Kane, D. (2012). *Modelling Single-name and Multi-name Credit Derivatives*. (D. O’Kane, Ed.). Hoboken, NJ, USA: John Wiley & Sons, Inc.

Patton, A. J. (2006). Modelling Asymmetric Exchange Rate Dependence. *International Economic Review*, *47*(2), 527–556.

Patton, A. J. (2009). Copula–Based Models for Financial Time Series. In *Handbook of Financial Time Series* (pp. 767–785). Berlin, Heidelberg: Springer Berlin Heidelberg.

Poon, S.-H., Rockinger Michael, & Tawn, J. (2004). Extreme Values, Regular Variation and Point Processes. *The Review of Financial Studies*, *17*(2), 581–610.

RiskMatrics Group. (1997). *CreditMetrics TM*. Retrieved from https://www.msci.com/documents/10199/93396227-d449-4229-9143-24a94dab122f

Rosenberg, J. V. (2003). Non-Parametric Pricing of Multivariate Contingent Claims. *The Journal of Derivatives*, *10*(3), 9–26.

Rosenberg, J. V., & Schuermann, T. (2006). A general approach to integrated risk management with skewed, fat-tailed risks. *Journal of Financial Economics*, *79*(3), 569–614.

Schweizer, B. (1991). Thirty Years of Copulas. In *Advances in Probability Distributions with Given Marginals* (pp. 13–50). Dordrecht: Springer Netherlands.

Sklar, M. (1959). *Fonctions de Répartition À N Dimensions Et Leurs Marges - M. Sklar - Google Books*. (229-231 Publ. Inst. Statist. Univ. Paris 8, Ed.).

The Economist. (2014). Banking without banks - Peer-to-peer lending. Retrieved from https://www.economist.com/news/finance-and-economics/21597932-offering-both-borrowers-and-lenders-better-deal-websites-put-two

Trivedi, P. K., & Zimmer, D. M. (2007). *Copula modeling : an introduction for practitioners*.

University of Oxford. (2016). *The Gaussian Copula and the Financial Crisis: A Recipe for Disaster or Cooking the Books?*

Vasicek, O. A. (1987). *Probability of Loss on Loan Portfolio*.

Ward, L. S., & Lee, D. H. (2002). *Practical Application of the Risk-Adjusted Return on Capital Framework*.

Zhang, B., Ziegler, T., Garvey, K., Ridler, S., Burton John, & Yerolemou, N. (2017). *Entrenching Innovation The 4th UK Alternative Finance Industry Report*.

1. <https://www.bankingtech.com/2018/11/the-emerging-landscape-of-fintech-regulation/> [↑](#footnote-ref-1)
2. <https://www.fundingcircle.com/uk/about-us/> [↑](#footnote-ref-2)
3. <https://www.fundingcircle.com/uk/about-us/> [↑](#footnote-ref-3)