Development and calibration of a multi-factor portfolio credit risk model

Mehmet Arkin (11604808), Sebastiaan Kruize (11673028) and Loek van Steijn (11811048)

February 2, 2022

COURSE Advanced Topics in Computational Finance

VU VRIJE UNIVERSITEIT AMSTERDAM | UNIVERSITEIT VAN AMSTERDAM

Abstract

Credit risk is a form of financial risk which concerns itself with the risk of default by a borrower that fails to pay their debt. For financial institutions like banks it is important to know about possible unexpected losses in their portfolio. In this work we have developed and calibrated a multi-factor Merton model so as to estimate portfolio credit risk. For this model we used systematic factors pertaining to region and industry sectors, constructed three different toy portfolios, and calibrated the model using market return data. As a measure of risk, we have estimated both Value-at-Risk and Expected Shortfall for each portfolio, using Monte Carlo simulations of potential loss events. Additionally, apart from the standard Gaussian copula that is often either implicitly or explicitly assumed in these models, we have implemented a version of this model that is based on the student-t copula. We have experimented with different levels of degrees of freedom, which show vastly different results for both VaR and ES when compared to the traditional Gaussian assumption. As it is hard to tell which copula reflects reality best, we conclude that risk managers should account for which copula better describes their portfolios best and estimate risk accordingly, so as to not underestimate potential portfolio loss.

Contents

1 Introduction

Credit risk is a form of financial risk which concerns itself with the risk of default by a borrower that fails to pay their debt. For financial institutions like banks it is important to know about possible unexpected losses in their portfolio. Adhering to the new market risk framework that was introduced by the Basel Committee on Banking Supervision in 2009, known as Basel 2.5, banks have to hold extra capital which is reserved for unexpected losses due to default and other risk charges in their portfolio so the financial institution remains solvent. How likely it is that a borrower defaults on a payment is determined by their creditworthiness (1), which is assigned by credit scoring agencies like Experian, TransUnion and Equifax. Another risk is rating migration risk where the creditworthiness of a borrower decreases and subsequently increases their probability of default. These risks are modeled by Default Risk Charge models (DRC) which satisfy regulation by the revised framework of the fundamental review of the trading book (FRTB) (2). They require that the default risk of the portfolio is measured independently of the market risk and that the model is simulated using two different types of systematic factors that contribute to the default behaviour of all obligors in a portfolio which usually contains some form of correlation. The correlation must be determined from data like equity prices spanning over a minimum period of 10 years for a sufficient confidence level and has at least one period of stress.

A well known model for credit risk is the Merton factor model (3) which, explicitly or implicitly, uses a copula to draw statistics for the systematic risk factors to include correlation and simulates a distribution of the portfolio loss due to obligor defaults based on their creditworthiness. The Gaussian copula is typically used but the European Central Bank (ECB) considers the Student-t copula with 8 degrees of freedom a suitable choice for the systematic risk factors (4). It also revers to the use of Student-t copulas where the degrees of freedom are calibrated to market data. A factor model decomposes economic variables into latent factors which are common to all variables and describes the co-movement of these variables. This is very useful for high dimensional data sets like portfolio's with many obligors because it reduces the dimensionality of the portfolio which improves, among other things, analysis like volatility forecasting or credit risk analysis if applied correctly. Calculating the correlation matrix of a large portfolio is computationally heavy which is not the case when the portfolio is described by a few factors. The dimensionality reduction is especially useful for credit risk models because this co-movement described by the factors ultimately sheds light on their default behaviour. The behaviour of obligors can be characterized by a number of factors including but not limited to the Global-, Industry- or Region factors. What the optimal number and choice of factors is has already been researched by I. Anagnostou *et al.* (5) and we have chosen to limit our factor model to two factors as by regulation of the Basel committee, which are the Industry sector factor and the Region factor.

An important note is the assumption that the default dependence between obligors, which is ultimately captured by our two factors, was Gaussian. This might have had a critical role in the credit crisis of 2008 so it is interesting to look at the influence of Copulas on the risk charge. There has been a lot of critique on the Gaussian assumption regarding the credit crisis (6), explaining that the tails of the Gaussian copula are too thin. The Gaussian assumption would therefore underestimate risk as extreme scenarios where multiple obligors default at once have a very low probability.

In this report we use the DRC regulations and the copula suggestions from the ECB to set up a multi-factor credit risk model with the Student-t copula where we analyse different degrees of freedom and compare them with the Gaussian copula. In addition, we show the influence of these copula choices on three portfolios with different concentrations to analyse concentration risk.

2 Method

2.1 Merton Model

The Merton Factor Model is frequently used to determine the value at risk (VaR) and estimated shortfall (ES) on credit portfolios. Its strength lies in the ease of reducing the dimensionality of the portfolios individual obligors, which especially for credit portfolios can very high, and unveiling the underlying correlations to certain systematic factors (7).
If we consider a credit portfolio containing N obligors we can define the potential loss L of the portfolio to be

$$L = \sum_{i=1}^{N} q_i e_i I_i \tag{1}$$

where q_i is the loss given default and e_i the exposure at default of obligor i. For the sake of simplicity we assume that the portfolio holder is exposed equally to all obligors and that the portfolio holder loses the entire exposure when a default occurs, this means we can set $q_i = 1$ and $e_i = \frac{1}{N}$. I_i is an indicator function which is 1 when a default occurs and zero otherwise. Thus, the model solely relies on determining the outcome of the indicator function. A default occurs when the creditworthiness of an obligor drops below the obligor specific default threshold d_i. The credit worthiness of an obligor can be seen as a random variable that can be expressed in terms of the systematic factors $\tilde{F}_i$, idiosyncratic factors ϵ_i and the sensitivity β_i.

$$X_i = \sqrt{\beta_i}\tilde{F}_i + \sqrt{1-\beta_i}\epsilon_i \tag{2}$$

Hereby assuming that the systematic factors $\tilde{F}_i$ are of the form $\tilde{F}_i = \vec{\alpha}_i^T \vec{F}$ where $\vec{F}$ is a vector of systematic factors so that $\vec{F} \sim N(0,\Omega)$ with Ω the correlation matrix of the systematic factors, and $\vec{\alpha}$ a vector containing the loading factors which satisfy $\vec{\alpha}^T \Omega \vec{\alpha} = 1$.
Combining everything we obtain

$$L = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{[-\infty, d_i]} \left(\sqrt{\beta_i} \tilde{F}_i + \sqrt{1-\beta_i} \epsilon_i \right) \tag{3}$$

and what remains is estimating the default thresholds for each obligor, the sensitivity factors β_i and the factor loadings α_i to which the following sections are dedicated.

2.2 Factor Calibration

As per FRBT guidelines we collect standardized non overlapping monthly log-returns of each obligor over a period of 10 years that includes at least one 'stressed' period. For this we choose the period 2011-2021 which contains the *Covid-19* pandemic as the stressed period. The Merton model assumes a liquidity horizon of a one year period which means that correlation should be measured in equal intervals which would ultimately mean that we are required to use the non overlapping yearly log-returns, however this would result in too few data points for calibration purposes. Taking the monthly log-returns instead is seen as a questionable but feasible compromise by Wilkens et al. (8) and Anagnostou et al. (5).
For this model we consider the frequently used systematic factors region and industry (7). To express the individual obligors in terms of these systematic factors we use linear regression

$$R_i = \gamma_r R_r + \gamma_s R_s \tag{4}$$

here the R_i is the standardized monthly log-return of obligor i, R_r and R_s the standardized orthogonal monthly log-returns of the systematic factors for which we used the returns of region and industry orientated ETFs respectively (see appendix for specifications) and γ_r and γ_s are the resulting regression coefficients. The sensitivity β_i and factor loadings $\vec{\alpha}_i$ used in expressing the obligors creditworthiness follow directly from the coefficient of determination R^2 and the normalized regression coefficients respectively.

2.3 Toy portfolios

In order to test the model we set up three toy portfolios of which the amount of correlation between the obligors contained in the portfolios seems to vary at first glace. Figure 1 shows the decomposition of our three toy portfolios sorted by the aforementioned systematic factors on which the Merton factor model will be based. All three portfolios contain 30 corporate obligors

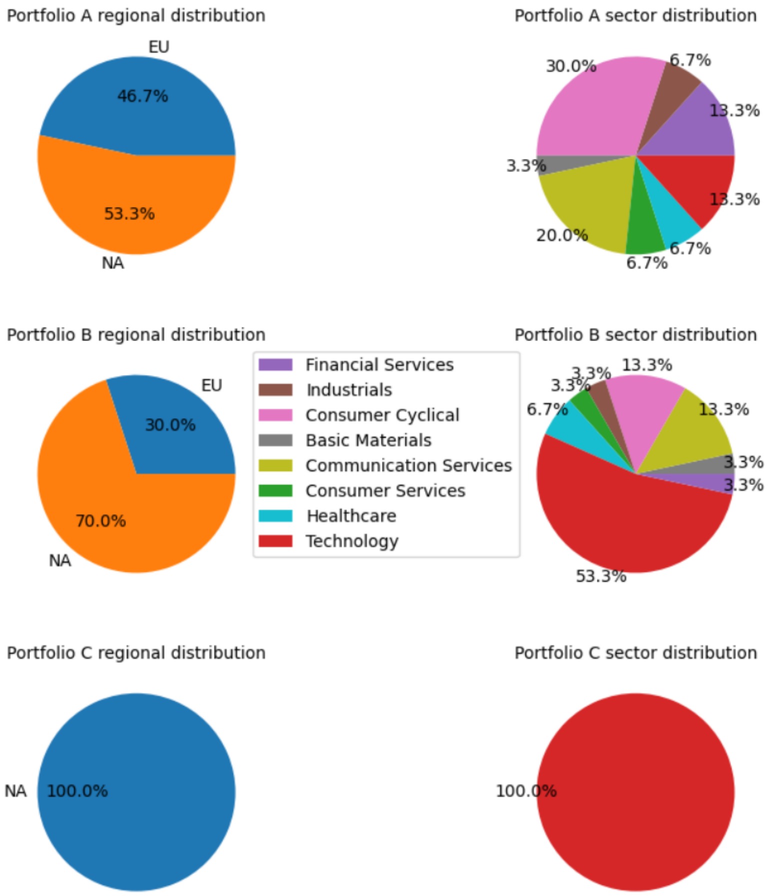

Figure 1: Decomposition of the three toy portfolios based on the systematic factors for the Merton factor model. The left hand side shows the regional distribution and the right hand side the industry distribution of the obligors.

each, which can be found in the Appendix. Banks and other credit procuring instances usually have portfolios far larger than this but to test the Merton factor model this will suffice. Portfolio A contains a mixture of European and North American corporations consisting from a wide selection of industries, Portfolio B is more concentrated towards North American Technology corporations while Portfolio C consists solely of North American Technology corporations.

As the model aims to simulate the likelihood of one or multiple obligors to default we must know what the probability of this event is. While banks and the like would normally infer this from their own accumulated data, we will base it on the probability of default (PD) per credit rating as issued by S&P Global Ratings (9), this data is based on annual default rates from 1981 to 2020. Table 1 shows the PD per credit rating as well as the composition within our toy portfolios where the last two columns show us the average and standard deviation of the PD from each portfolio. The Merton model simulates states where the obligors creditworthiness X_i

	AA	A	BBB	BB	Average PD	Std PD
PD	0.0141%	0.0544%	0.1997%	0.8538%		
Portfolio A	3	11	14	2	0.17147%	0.1973%
Portfolio B	1	11	16	2	0.1838%	0.1928%
Portfolio C	0	10	17	3	0.2167%	0.2226%

Table 1: The probability of default per credit rating as issued by S&P Global Ratings and the credit rating composition of each of the toy portfolios.

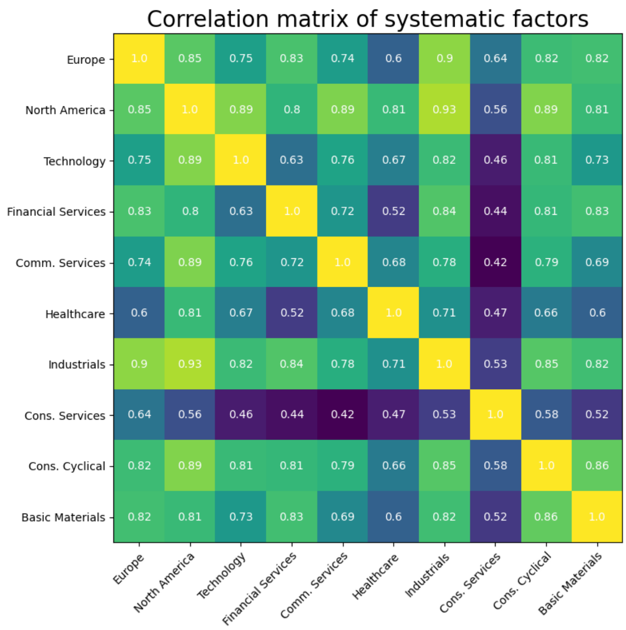

Figure 2: Correlation matrix of systematic factors.

drops below the default threshold d_i. The systematic factors are based on normally distributed random variables thus the creditworthiness will follow the same normal distribution as a result the default threshold can be determined by taking the cumulative normal distribution at this point

$$d_i = \Phi^{-1}(PD_i) \tag{5}$$

here $\Phi(\cdot)$ is the cumulative normal distribution and PD_i the probability of default of obligor i.

2.4 Simulation

First we create a correlation matrix of all the systematic factors as is shown in Figure 2. For the basic model we then fit a Gaussian copula (see next section) to the correlation matrix. Next we perform the aforementioned calibration step to express the individual obligor in terms of the systematic factors and thus reducing the dimensionality of the portfolio and to acquire the sensitivity β_i and factor loadings α_i for each obligor. We then use a Monte Carlo algorithm to simulate the portfolio loss where in each trial we sample the Gaussian copula and for each obligor determine the creditworthiness according to Equation 2 with $\tilde{F}_i = \vec{\alpha}_i^T \vec{F}$ and the idiosyncratic risk factor $\epsilon_i \sim N(0,1)$, and followed by determining if the obligors creditworthiness drops below their default threshold. Lastly, we add all of the portfolio loss together according to Equation 3. We repeat this process 10^7 times from which we can determine the VaR and ES. The VaR of a portfolio with loss L at a confidence interval ω is defined by the smallest number l such that the probability that the incurred loss is higher than l does not exceed $1-\omega$ over a given period, in our case 1 year. ES is then the expected loss in the cases where the VaR is exceeded. In our model the VaR can be calculated by adding each simulated loss to a list, ordering the list in ascending order and taking the loss at the $10^7 \times \omega^{th}$ index. The ES can be determined by taking the mean of the losses high than the $10^7 \times \omega^{th}$ index. A pseudo example can be found in Algorithm 1.

Algorithm 1 Monte Carlo simulation for determining credit risk.

Require: fitted copula, sensitivity β, loading factors α, default thresholds d

```
default list as empty list
for j → 10^7 do
    Loss = 0
    F⃗ = copula.random(1) and scale from uniform to normal
    for i → N do
        X_i = √β_i × (α⃗_i^T · F⃗) + √(1 − β_i) × random.normal(0,1)
        if X_i < d_i then
            L_i = 1/N
        else
            L_i = 0
        end if
        Loss = Loss + L_i
    end for
    add Loss to default list
end for
order default list (ascending)
VaR = default list[N × ω]
ES = mean( default list[N × ω :])
```

2.5 Copulas

The joint distribution function of random vectors containing systematic factors consists of the behaviour of individual factors but also the behaviour of the dependence between these factors. Applying copulas to such random vectors can extract the dependence part of these factors which in turn results in a better understanding of the dependence structure. The marginal behaviour of individual obligors is often properly understood, introducing a copula allows us to combine these marginal models with different dependence structures to investigate the loss/risk sensitivity better.

These properties come forth through Sklar's theorem which states; Let F be a d-dimensional distribution function with margins F_1 to F_d, then there exists a copula C which is unique if the margins are continuous i.e.

$$F(x_1,...,x_d) = C(F_1(x_1),...,F_d(x_d)). \tag{6}$$

Furthermore it shows that all multi dimensional distribution functions contain copulas and that copulas may be used to construct a multivariate from univariate distribution functions F_1 to F_d.

For our model we focus on parametric copulas, namely the Gaussian and, for fatter-tail distributions, the Student-t copulas (10). The Gaussian copula can be defined as

$$C_\Omega^{\text{Gauss}}(\mathbf{x}) = \Phi_\Omega\left(\Phi^{-1}(x_1),\dots,\Phi^{-1}(x_d)\right), \tag{7}$$

where $\Phi(\cdot)$ is the univariate cumulative distribution function and $\Phi_\Omega(\cdot)$ the joint cumulative distribution function under correlation matrix Ω. The Student-t copula is then defined as

$$C_{\nu,\Omega}^t(\mathbf{x}) = \mathbf{t}_{\nu,\Omega}\left(t_\nu^{-1}(x_1),\dots,t_\nu^{-1}(x_d)\right) \tag{8}$$

where t_ν is the univariate cumulative t-distribution function and $t_{\nu,\Omega}$ the joint cumulative t-distribution function with ν degrees of freedom.

We can fit the aforementioned copulas to our data in several different ways, the simplest is via the method-of-moments using rank correlations. For the Gaussian copula we can take the *Spearman's* rank correlation coefficient that is given by, assuming bivariate data

$$r_{ij}^S = \frac{12}{n(n^2-1)}\sum_{t=1}^{n}(rank(X_{t,i}) - \frac{1}{2}(n+1))(rank(X_{t,j}) - \frac{1}{2}(n+1)) \tag{9}$$

where $rank(X_{t,i})$ is the position of $X_{t,i}$ in an ordered list. Calibrating this using the *Spearman's rho* coefficient

$$\rho_S(X_i, X_j) = \frac{6}{\pi} \arcsin\left(\frac{1}{2}\rho_{ij}\right) \approx \rho_{ij} \tag{10}$$

we see that we can estimate the correlation matrix Ω by taking the pairwise $r^S_{i,j}$ values at the respective indices. Similarly, for the Student-t copula we can use the same method but using *Kendall's* rank correlation coefficient

$$r^\tau_{ij} = \binom{n}{2}^{-1} \sum_{1\le t<s\le n} sign((X_{t,i} - X_{s,i})(X_{t,j} - X_{s,j})) \tag{11}$$

Again, we can calibrate this using the *Kendall's tau* coefficient

$$\rho_\tau(X_i, X_j) = \frac{2}{\pi} \arcsin \rho_{ij} \tag{12}$$

from where it can be derived that the correlation matrix Ω can be estimated by taking $\sin(\frac{\pi}{2} r^\tau_{ij})$ at the respective indices. However several (pseudo) maximum likelihood estimations can also be used to fit copulas, such as is used in most *Python* packages including *Copulae* which was used during this project. Calibration of the resulting correlation matrix follows the same procedure.

In practise, fitting solely a Gaussian copula to the data is sufficient as this can be scaled to form a Student-t copula. First we sample a Chi squared distribution with ν degrees of freedom and call it χ, then we define scaling weight $W = \sqrt{\frac{\nu}{\chi}}$. We can then simulate a Student-t copula by scaling Equation 2 to

$$X_i = \sqrt{\beta_i}\tilde{F}_i W + \sqrt{1-\beta_i}\epsilon_i W \tag{13}$$

but now with a default threshold of $d_i = t_\nu^{-1}(PD_i)$ with $t_\nu(\cdot)$ the cumulative Student-t distribution function with ν degrees of freedom (7; 11).

3 Results

The main result of the experiment is shown in Table 2, and for a more visual representation, we show the same results in Figures 3 and 4. In addition to the main results, we also show the loss distributions under each copula in Figures 5-10 to get a more detailed overview. First, we see that using a completely uncorrelated copula, the risk measures are, as we would expect, much lower than other copulas that incorporate a dependence structure between obligors. We

	α	**Uncorr**	**Gaussian**	**t-100**	**t-20**	**t-8**	**t-3**
Port A	0.990	3.3 (3.73)	3.3 (7.13)	3.3 (7.48)	3.3 (8.66)	3.3 (10.51)	3.3 (13.32)
	0.995	3.3 (4.13)	6.7 (9.48)	6.7 (9.95)	6.7 (11.71)	6.7 (14.94)	10 (20.52)
	0.999	6.7 (6.73)	10 (16.56)	13.3 (17.65)	13.3 (22.02)	20 (29.67)	26.7 (43.02)
Port B	0.990	3.3 (3.80)	3.3 (7.06)	3.3 (7.41)	3.3 (8.80)	3.3 (10.90)	3.3 (14.14)
	0.995	3.3 (4.27)	6.7 (9.12)	6.7 (9.58)	6.7 (11.54)	6.7 (15.15)	10 (20.34)
	0.999	6.7 (6.74)	10 (15.12)	10 (16.42)	13.3 (20.72)	20 (28.52)	30 (43.20)
Port C	0.990	3.3 (3.98)	3.3 (8.46)	3.3 (8.84)	6.7 (10.16)	6.7 (12.25)	6.7 (15.94)
	0.995	3.3 (4.62)	6.7 (10.75)	6.7 (11.28)	6.7 (13.65)	10 (17.41)	10 (24.11)
	0.999	6.7 (6.78)	13.3 (18.65)	13.3 (19.82)	16.7 (24.88)	23.3 (33.32)	33.3 (48.25)

Table 2: Risk measure estimate results for all portfolios and copulas. Displayed is the confidence level α, the copula used, and the corresponding VaR (ES) estimates in terms of the **percentage** of the portfolio lost.

also see this in Figure 5, where we see at most 4 defaults happening (out of a portfolio of 30 obligors), while the Gaussian copula in Figure 6 already shows over 15 defaults possibly happening. What is clear from the results is that as copulas with stronger tail dependence are incorporated, VaR and ES estimates significantly increase, as we get more dependence in the tails and thus more joint default events. As we start using the t-copula and decrease the amount of degrees of freedom, we can see in Figures 7-10 that the loss distributions obtained from the simulations start to 'creep up' to having some default events where almost every obligor defaults simultaneously, meaning a very heavy tail dependence structure. Clearly, this means that estimating credit loss is quite sensitive to the underlying dependence structure assumptions.

Another thing to notice here is that portfolio B seems to have lower risk measures than portfolio A under some copulas, while it is both more concentrated and has a higher average probability of default. This is an interesting result, as when we compare this to portfolio C, which is fully concentrated in the technology sector, the risk measures are higher, and this is what we would expect. This result in especially visible in Figures 6-10, where we see that the middle portfolio, B, has more occurrences at the lower default levels, but fewer occurrences further down the tail, when compared to portfolio A.

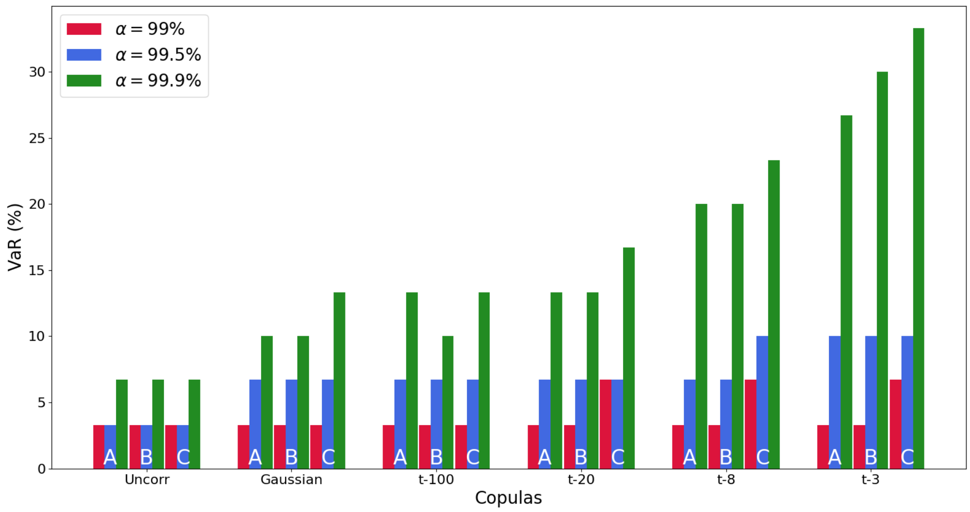

Figure 3: VaR estimates for all portfolios and copulas.

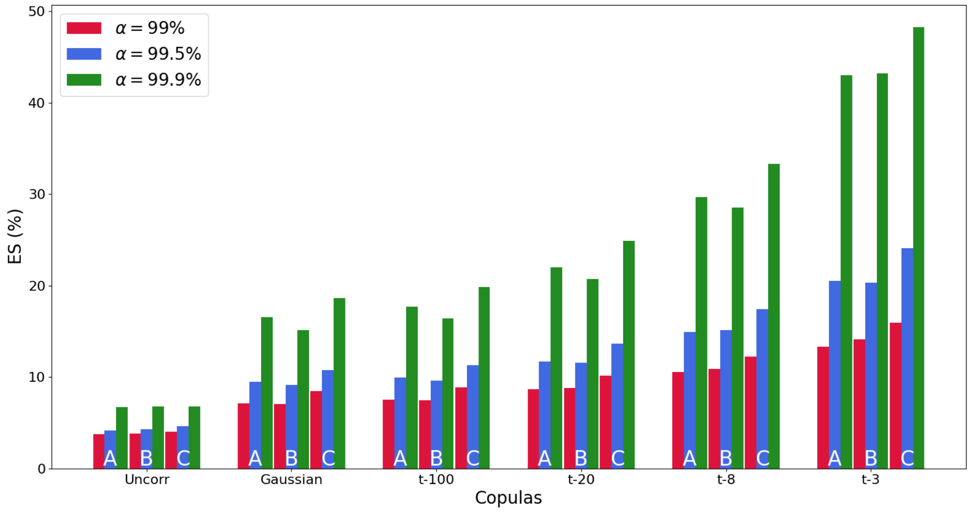

Figure 4: ES estimates for all portfolios and copulas.

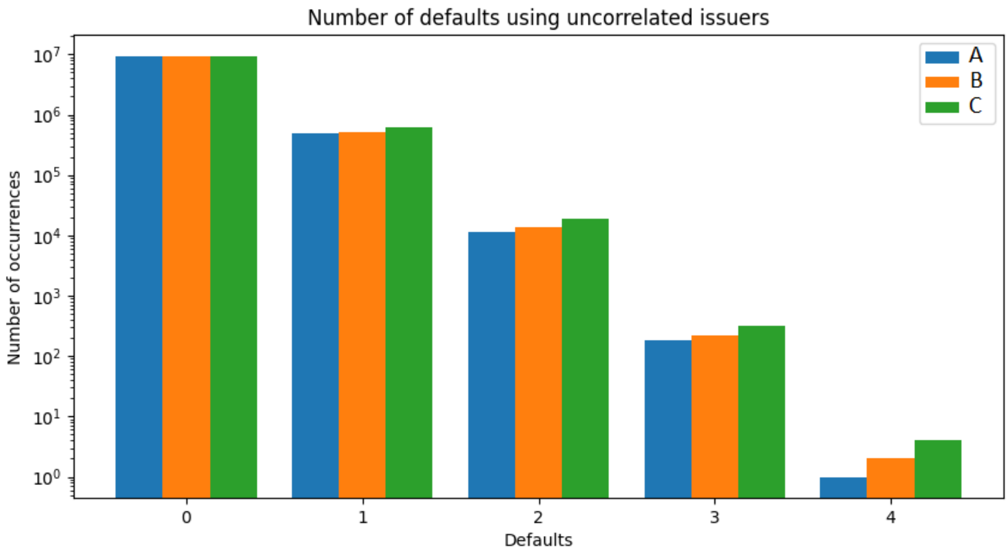

Figure 5: Loss distribution after 10 million Monte Carlo trials without a dependence structure between obligors. Portfolio contains 30 obligors.

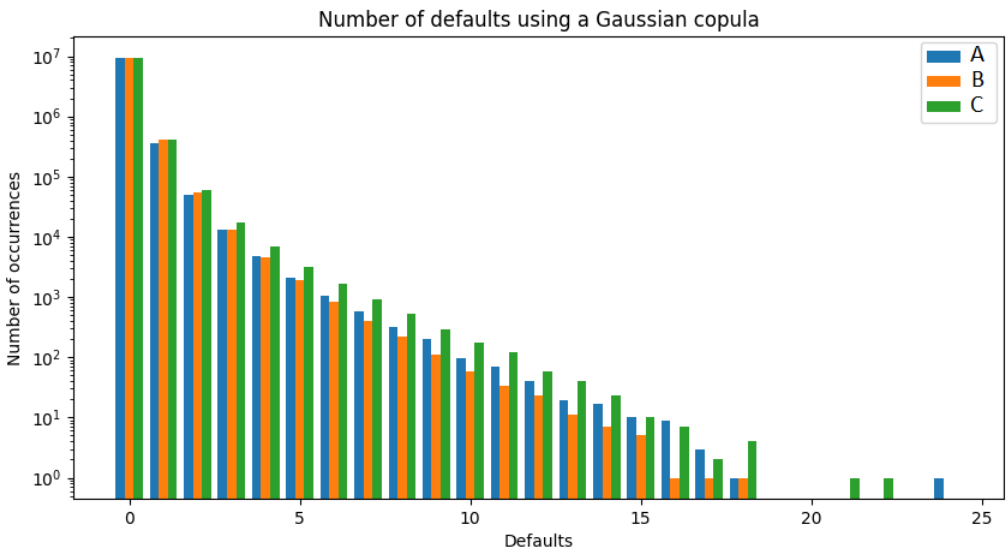

Figure 6: Loss distribution after 10 million Monte Carlo trials using the Gaussian copula to model the dependence structure between obligors. Portfolio contains 30 obligors.

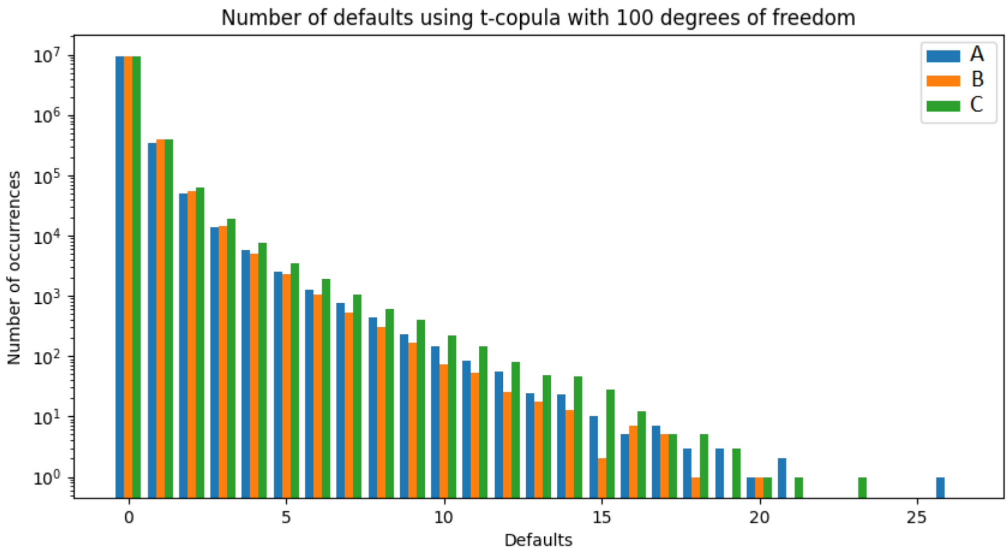

Figure 7: Loss distribution after 10 million Monte Carlo trials using the Student-t copula with 100 degrees of freedom to model the dependence structure between obligors. Portfolio contains 30 obligors.

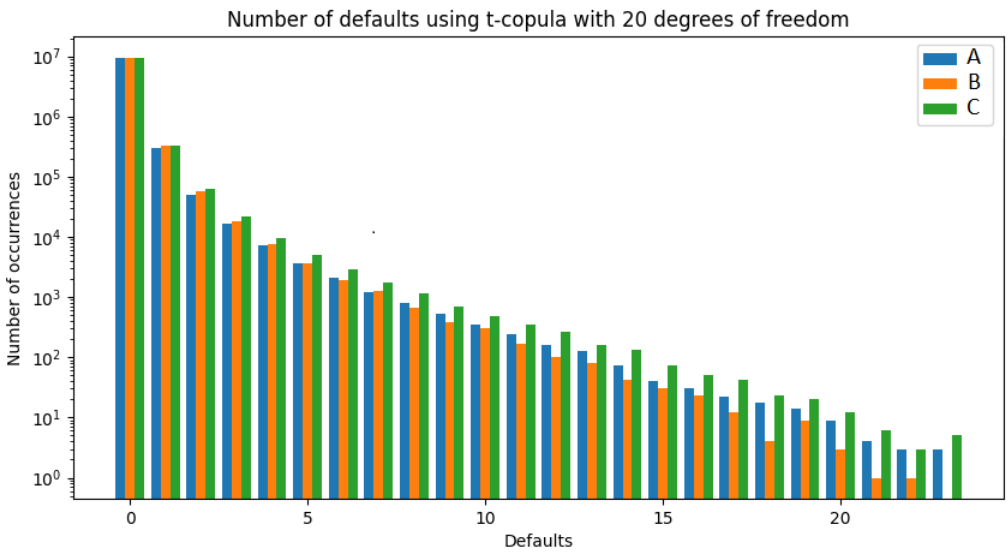

Figure 8: Loss distribution after 10 million Monte Carlo trials using the Student-t copula with 20 degrees of freedom to model the dependence structure between obligors. Portfolio contains 30 obligors.

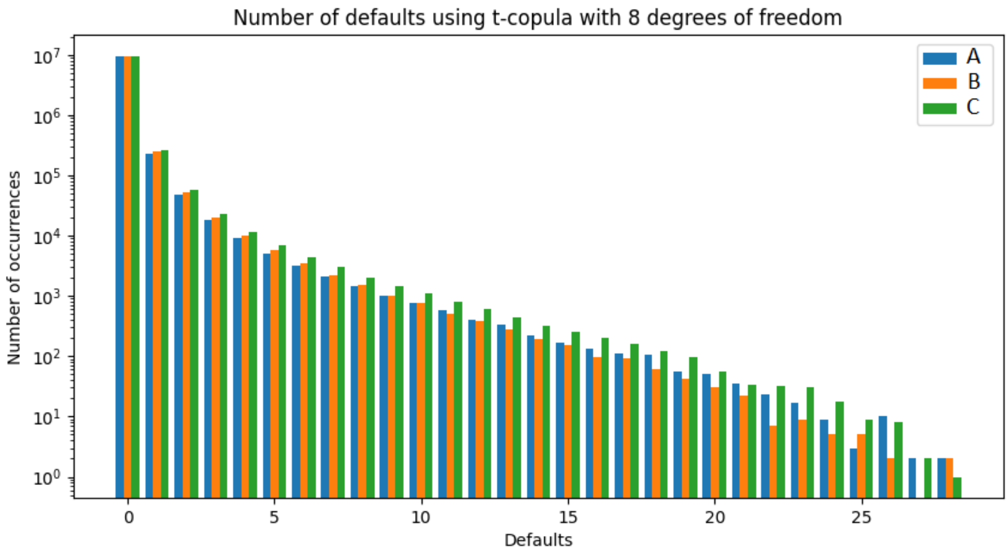

Figure 9: Loss distribution after 10 million Monte Carlo trials using the Student-t copula with 8 degrees of freedom to model the dependence structure between obligors. Portfolio contains 30 obligors.

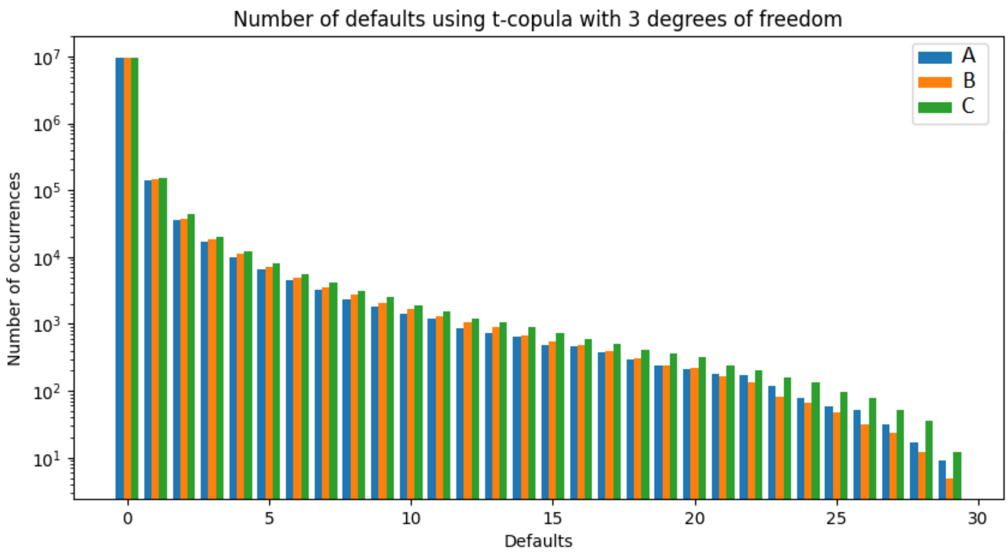

Figure 10: Loss distribution after 10 million Monte Carlo trials using the Student-t copula with 3 degrees of freedom to model the dependence structure between obligors. Portfolio contains 30 obligors.

4 Discussion

The results indicate that using copulas other than the usual Gaussian copula can make a significant difference in risk measure. Especially when a heavy-tailed copula such as the student-t copula with three degrees of freedom is used, we see significantly larger Value-at-Risk and Expected Shortfall estimates at the 99.9% confidence level. It is also clearly visible that as one increases the amount of degrees of freedom using this copula, the loss distributions start to resemble the Gaussian case, and when 100 degrees of freedom are used, the distribution is very similar the Gaussian copula, at least where VaR and ES estimates are concerned, and we can expect the two meeting exactly if one were to increase the degrees of freedom even further. However, it is surprising that even when using a fairly light-tailed student-t copula with twenty degrees of freedom, the risk measures are still noticeably more conservative than the Gaussian model. As a key message, we see that using eight degrees of freedom, as per the recommendation of the European Central Bank, shows consistently higher risk measures than the Gaussian model, especially at the 99.9% confidence level. This suggests that a credit risky portfolio may indeed be subjected to more potential loss than risk managers would anticipate, and thus the choice of copula is an important facet that must be taken into account.

An additional result we show is that all copulas seem to handle concentration risk in a similar fashion, as we see in Table 2 that the ratios between risk measures across all portfolios are quite similar for each copula.

It is interesting to mention, however, that Portfolio B, which is the half-tech portfolio, seems to consistently have a higher number of occurrences at the lower default levels, but a *lower* number of occurrences at the higher default levels, when compared with the more spread Portfolio A if we look at Figures 5-10. This seems to go against intuition, as one could expect that high obligor correlation in the portfolio would result in high concentration risk and thus higher risk measure estimates, as with Portfolio C, but this does not seem to be the case. It is especially interesting since the average probability of default in Portfolio B is higher than that in Portfolio A, which means this anomaly is not necessarily explained by simply having a different average probability of default. At this point, it is difficult to say whether this is simply a quirk of the model, portfolio, or otherwise, and this could be looked into further.

Concretely, in this work we have shown that using the appropriate copula as a base assumption for a credit risk model can lead to significantly higher risk measure estimates. In this regard, it seems like the critique on the Gaussian assumption regarding the credit crisis of 2008 is not out of place, as we have shown that it indeed seems to comparatively underestimate tail risk.

Of course, the results shown in this work are merely indicative and should be treated as such. The datasets used for model calibration have often not been complete, and have contained less data than one would typically desire for adequate calibration. Also, the results seem to suggest that they are quite portfolio dependent, as we have found some counter-intuitive results, which makes the choice of copula quite case-specific, as is perhaps what one might expect. Additionally, even though we have shown that the use of some copulas results in a more conservative risk measure, we cannot say anything about the value of such a result, as it is difficult to say which copula reflects reality best. As such, we cannot say that using one copula over another is better or worse, for the time being.

Concerning additional work that could be done in the future, there are a few things that spring to mind. First, a big problem with doing these types of Monte Carlo simulations is that it can take an extraordinarily long time to attain adequate results. This is especially relevant for the work done here, as what we are really interested in are low-probability, tail-end occurrences that only happen in a very minor fraction of all simulations. For example, 99.9% VaR estimates are based on events that happen once per thousand simulations. Long story short, this means that in order to get a representative and statistically significant result, a lot of these simulations must be done, and this can get extremely computationally expensive. A logical conclusion here is that variance reduction techniques, such as importance sampling, would make for much shorter computation times, and as such it is our first recommendation that importance sampling is added to the model, which is a technique that is not uncommon for portfolio credit risk models (12; 13). The essence of importance sampling is that one samples from a different distribution than the distribution of interest. In our context this would imply sampling from a distribution that represents the low-probability tail end of the distribution of interest. This of course results in a sampling bias, as just sampling from the tail for low-probability occurrences is not realistic. One accounts for this by assigning weights to occurrences which are given by the likelihood ratio (14). More applicably speaking, for our default model, one could scale up the systematic risk factors in order to *force* occurrences in the region of interest, and then assign the correct weights. This method is known as *shifting the factor mean* (15). Another method is *exponential twisting*, where the default probabilities are altered and, again, corrected by the likelihood ratio. The caveat here is that importance sampling can be tricky to implement because of the dependence structures between obligors, and is thus also copula-dependent, which further complicates things. Finding a suitable scale factor then could become intricate.

Secondly, it would likely be a fruitful endeavor to look into different methods of model calibration. In the calibrations done in this work, finding sufficiently rich datasets was not always accomplished, and often many data points were missing, up to the point where half of the desired data was missing at times, having for example five years of data instead of the desired ten years or more. Having a few years of data is not always a problem, but in this specific case, where we look into expected yearly losses, it means that it is insufficient. Further, concessions have already been done in this case, using monthly return data instead of yearly, since regression done on only a handful of data points would prove inadequate. However, the validity of this concession could be contested (8). As such, different methods of calibration would be advisable, such as calibration through for example CDS spread data of obligors (16; 5), with the side note that these types of data are not always complete or available either, and in those cases other sources of data or calibration methods should be looked into further.

Lastly, there is no reason why copulas outside of the Gaussian and student-t shouldn't be tried, and it might prove to be interesting to look at the effects of copulas such as the Gumbel copula. In some cases, the dependence structure of returns in the portfolio may be asymmetrical, for example there may be higher dependence in downward market movement as opposed to upward movements, meaning that an asymmetrical copula could be a better fit, and similar work has indeed already been done (17). It would therefore be an interesting addition to add different types of copulas to the model, which can explore vastly different dependence structures. Not only this, but since in this work only industry and region systematic factors were used, it could be interesting to look at not only the impact of these other copulas, but also the impact of these copulas where different systematic factors are used.

5 Bibliography

[1] E. I. Altman, "An emerging market credit scoring system for corporate bonds," *Emerging markets review*, vol. 6, no. 4, pp. 311–323, 2005.

[2] Basel_Committee_on_Banking_Supervision, "Fundamental review of the trading book: A revised market risk framework," 2013.

[3] S. T. Bharath and T. Shumway, "Forecasting default with the merton distance to default model," *The Review of Financial Studies*, vol. 21, no. 3, pp. 1339–1369, 2008.

[4] European_Central_Bank, "Ecb guide to internal models," 2019.

[5] I. Anagnostou, T. Squartini, D. Kandhai, and D. Garlaschelli, "Uncovering the mesoscale structure of the credit default swap market to improve portfolio risk modelling," *Quantitative Finance*, pp. 1–18, 2021.

[6] D. MacKenzie and T. Spears, "The formula that killed wall street? the gaussian copula and the material cultures of modelling," *School of Social and Political Science, University of Edinburgh*, 2012.

[7] A. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management: concepts, techniques and tools*. Princeton University Press, 2015.

[8] S. Wilkens and M. Predescu, "Default risk charge: Modeling framework for the 'Basel' risk measure," *Journal of Risk*, vol. 19, no. 4, 2017.

[9] N. Kraemer, "2020 annual global corporate default and rating transition study," 2021.

[10] A. Lourme and F. Maurer, "Testing the gaussian and student's t copulas in a risk management framework," *Economic Modelling*, vol. 67, pp. 203–214, 2017.

[11] M. Haugh, "An introduction to copulas," 2016.

[12] P. Glasserman and J. Li, "Importance sampling for portfolio credit risk," *Management science*, vol. 51, no. 11, pp. 1643–1656, 2005.

[13] R. Chen, Z. Wang, and L. Yu, "Importance sampling for credit portfolio risk with risk factors having t-copula," *International Journal of Information Technology & Decision Making*, vol. 16, no. 04, pp. 1101–1124, 2017.

[14] B. Arouna, "Adaptative monte carlo method, a variance reduction technique," 2004.

[15] P. Glasserman and J. Li, "Importance sampling for portfolio credit risk," *Management science*, vol. 51, no. 11, pp. 1643–1656, 2005.

[16] I. Anagnostou, S. Sourabh, and D. Kandhai, "Incorporating contagion in portfolio credit risk models using network theory," *Complexity*, vol. 2018, 2018.

[17] M. Fischer and K. Jakob, "Copula-specific credit portfolio modeling," in *Innovations in Quantitative Risk Management*, pp. 129–145, Springer, 2015.

6 Appendix

Table 3 shows an overview of all the corporations included in the toy-portfolios. Table 4 shows an overview of all return data used to establish the systematic factors.

ABN AMRO Bank N.V.	Atlantia SpA	International Paper Company
ING Groep N.V.	adidas AG	Kellogg Company
Aegon N.V.	Kingfisher plc	McDonald's Corporation
Siemens Aktiengesellschaft	British American Tobacco p.l.c.	Electronic Arts Inc.
Bayerische Motoren Werke AG	Activision Blizzard, Inc.	Ford Motor Company
BASF SE	CVS Health Corporation	General Motors Company
Vodafone Group Plc	Apple Inc.	Broadcom Inc.
Renault SA	IBM Corporation	Intel Corporation
Deutsche Telekom AG	Abbott Laboratories	The Walt Disney Company
Telefónica, S.A.	Berkshire Hathaway Inc.	NIKE, Inc.
Adobe Inc.	Analog Devices, Inc.	Citrix Systems, Inc.
Advanced Micro Devices, Inc.	Autodesk, Inc.	Corning Incorporated
Fiserv, Inc.	NVIDIA Corporation	Trimble Inc.
FLIR Systems, Inc.	Oracle Corporation	VeriSign, Inc.
HP Inc.	PTC Inc.	Western Digital Corporation
Intuit Inc.	QUALCOMM Incorporated	Zebra Technologies Corporation
Lam Research Corporation	salesforce.com, inc.	Juniper Networks, Inc.
Motorola Solutions, Inc.	ServiceNow, Inc.	Texas Instruments Incorporated
NetApp, Inc.	Skyworks Solutions, Inc.	NortonLifeLock Inc.

Table 3: Overview of all obligors in the toy portfolios

Factor	Index
Europe	MSCI EUROPE
North America	MSCI NORTH AMERICA
Technology	MSCI WRLD/TECHNOLOGY
Financial Services	MSCI WRLD/FINANCIALS
Communication Services	MSCI WRLD/COMMUNICATIONS SVC
Health Care	MSCI WRLD/HEALTH CARE
Industrials	MSCI WRLD/INDUSTRIALS
Consumer Services	MSCI WRLD/CONSUMER SVC
Consumer Cyclical	MSCI WRLD/CONSUMER DISC
Basic Materials	MSCI WRLD/MATERIALS

Table 4: Indices used for systematic factors.