Inaccurate Dependence Measures in Credit Models for Non-Normal Variables

Fernando Moreira

Keele University, UK

Due to the assumption of normally distributed variables, conventional credit models have been criticized for not being able to identify possible extreme losses. As an alternative, some methods have incorporated non-normal variables in the estimation of the probability of default in loan portfolios and credit derivatives. One of the objectives of these methods is to express heavy tails of the distributions (which tends to better represent the reality of the credit market since economic and financial variables typically present more extreme occurrences than indicated by the normal distribution). However, as this paper shows, the derivation of some of these alternative models does not comply with all the assumptions implicit in the formula used to develop the models and this mistake results in misleading dependence measures. Our theoretical arguments are supported by simulations which show that, in terms of the calculation of regulatory capital for financial institutions, models for non-normal variables overestimate losses and this bias is substantial for high levels of confidence (up to 13 times higher than the losses observed in the simulated credit portfolio). We present some ideas to start solving this problem although the estimation of the dependence parameter is still an open question.

JEL classification: G29; G32;, C46; C49
Keywords: Credit risk, non-normal variables, dependence measure, inaccuracy, copulas

1. Introduction

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

To overcome such limitations, non-normal variables have been inserted in some credit risk models in order to capture higher proportion of extreme events (than captured in approaches based on the normal distribution). We show in this paper that, although these methods have improved one aspect of the previous credit models (i.e. the unrealistic assumption of normality for each variable considered), they have created a new problem regarding the dependence across the variables studied given that important conditions inherent in their statistical formulation have been neglected. This drawback has been ignored in the literature and some authors have applied the alternative approach without realizing its flaw.

Our simulations confirm that the dependence measures used in these alternative models do not correspond to the real dependence measure of the data when a distribution different from the normal was used to represent particular variables. As a consequence, estimates of losses in unfavorable scenarios are biased. We found that these models have the merit of avoiding the underestimation of losses in severe conditions but they result in highly overestimated losses which may be disadvantageous in some cases. Broadly speaking, the overestimation level decreases with the probability of default and increases with the confidence demanded.

Moreover, the aforementioned methods for non-normal variables focus on the marginal (univariate) distributions of the variables and do not pay enough attention to the dependence structure that links the variables and also affects the probability of unexpected events. Inserting such connection structure into the analysis seems to be a way to find more accurate dependence parameters and, consequently, to achieve more precise calculations of the probability of high losses.

In this context, the main contributions of this paper are threefold. Firstly, we warn about the misleading dependence measures used in these recently suggested credit models such that their users are aware of their limitations and potential biases. Secondly, we show the origin of this inconsistency and its impact on the estimation of credit losses in extremely unfavorable scenarios. Thirdly, we indicate directions to the development of more precise models by highlighting the importance of the dependence structure in the estimation of extreme losses. To our knowledge, none of these contributions have been discussed in the literature so far.

The paper proceeds as follows: in the next section, we explain how some credit models related to the estimation of the probability of default are derived by means of factor models when variables are assumed to be normally distributed. We also present some models suggested to deal with non-normal variables. In Section 3, we explain why the dependence measure employed in these alternative models is inaccurate and then we show the impact of this inaccuracy on the estimation of credit losses in severe conditions. Section 4 presents some ideas to start solving this problem. Section 5 concludes.

2. Credit Risk Models: Factor Models and the Assumption of Normality

2.1. Measuring joint probability of default via factor models

Structural credit models (originally proposed by Merton, 1974) consider that defaults happen when the return of obligors' assets (a latent variable) falls below a specific value (the amount needed to pay the outstanding debt). The probability of default (*PD*) is the probability of the obligors' asset returns falling below the threshold value.

If we are interested in estimating the dependence across defaults of different obligors, we can use factor models which assume that the correlation among defaults is driven by the debtors' latent variables (see, for instance, Crouhy et al., 2000 and Bluhm et al., 2002). These underlying variables are impacted by common (systematic) factors that affect all obligors and specific (idiosyncratic) factors that have effect only on the respective borrowers. The idiosyncratic factors are assumed to be independent from one another and therefore do not contribute to asset return correlations which are exclusively determined by the systematic factors.

We can simplify this model by considering that the asset returns of all borrowers are driven by only one common factor (the "economic status") and by assuming that those latent variables (the asset returns) can be expressed as a linear function of the common (systematic) factor and the specific (idiosyncratic) risk:

$$Y_i = \beta_1 X + \beta_2 \varepsilon_i \qquad 2.1$$

where Y_i is the latent variable of obligor i, X is the systematic factor, ε_i is the idiosyncratic factor for obligor i, and β_1 and β_2 are coefficients that indicate how much of the variation in Y_i is explained by X and ε_i respectively.

Some popular credit models (for example, CreditMetrics® and KMV®) adopt approaches based on factor models and assume that the latent variable (Y), the single systematic factor (X), and the specific factor (ε) are standardized normally distributed. Each idiosyncratic risk is supposed to be uncorrelated with the systematic risk and the specific risks of all other obligors. For simplicity, all pairs of asset returns (i and j) are considered to present the same correlation (ρ_{ij}). The correlation between the systematic factor and the asset return of each debtor is denoted ρ_{YX}.

Owen and Steck (1962) showed that equally correlated and jointly standard normal variables (in our case, the latent variables of two obligors i and j for example) may be expressed as a function of their correlation coefficient (ρ_{ij}) and another two standard normal variables (here, X and ε_i). Therefore, considering all assumptions of credit models mentioned above, the coefficients β_1 and β_2

in (2.1) are associated with ρ_{YX} and (2.1) becomes[1]:

$$Y_i = X\sqrt{\rho_{ij}} + \varepsilon_i\sqrt{1-\rho_{ij}} \tag{2.2}$$

where

$$\sqrt{\rho_{ij}} = \rho_{YX} \tag{2.3}$$

since the idiosyncratic risk is assumed to be independent and all the variables are standardized with mean 0 and variance 1 (see proof in Moreira, 2011, Appendix A).

Expression (2.2) does not hold for distributions other than the normal. So, the use of the correlation coefficient ρ_{ij} in a linear function to evaluate each latent variable Y is conditional on the normality of the variables involved (joint normality between Y_i and Y_j and univariate normality of X, ε_i, and ε_j).

We can use (2.2) to derive a formula to estimate the probability of default conditional on particular events or on particular economic levels (downturns, for instance). As said before, for each loan i, the probability of default is the likelihood that the latent variable Y_i becomes smaller than the cutoff y_c, that is, $PD = \Pr[Y_i < y_c]$. The probability of default, PD^*, when the economy X reaches the level x^*, is given by $PD^* = \Pr[Y_i < y_c \mid X = x^*]$. Using (2.2), we have:

$$PD^* = \Pr[X\sqrt{\rho_{ij}} + \varepsilon_i\sqrt{1-\rho_{ij}} < y_c \mid X = x^*]$$

Solving for ε_i and replacing X with x^*:

$$PD^* = \Pr\left[\varepsilon_i < \frac{y_c - \sqrt{\rho_{ij}}\,x^*}{\sqrt{1-\rho_{ij}}}\right]$$

Since ε_i is presumed to be normally distributed with mean 0 and variance 1, the previous equation turns into:

$$PD^* = \Phi\left(\frac{y_c - \sqrt{\rho_{ij}}\,x^*}{\sqrt{1-\rho_{ij}}}\right) \tag{2.4}$$

where Φ indicates the cdf (cumulative distribution function) of the standard normal distribution.

Given that Y_i is also normally distributed, $PD = \Phi(y_c)$ which implies that $y_c = \Phi^{-1}(PD)$, i.e. the cutoff of the latent variable below which default occurs is the inverse of the normal distribution, Φ^{-1}, evaluated at PD. The level of X when it is equal to x^* refers to the area below the point x^* in the X distribution. Denoting this area as A_X^*, we have $A_X^* = \Phi(x^*)$ and therefore $x^* = \Phi^{-1}(A_X^*)$. Thus, replacing y_c and x^* in (2.4), the probability of default conditional on the economic status x^* becomes:

$$PD^* = \Phi\left(\frac{\Phi^{-1}(PD) - \sqrt{\rho_{ij}}\,\Phi^{-1}(A_X^*)}{\sqrt{1-\rho_{ij}}}\right) \tag{2.5}$$

where A_X^* indicates the economic level. Additional details on this derivation can be found in Schönbucher (2000) and Perli and Nayda (2004).

(2.5) has a practical application, for example, in Basel Accords to determine the capital financial institutions should set aside to cover unexpected credit losses (with A_X^* = 0.001 and

[1] An expression equivalent to (2.2) is valid for Y_j (by replacing ε_i with ε_j).

$\Phi^{-1}(0.001) = -\Phi^{-1}(0.999)$ which represents a confidence of 99.9%, i.e. the potential credit losses when the economy reaches the 99.9th worst scenario)[2].

2.2. Drawbacks of the normality assumption

As seen above, the assumption of normality is essential for the derivation of credit models based on factor models but two aspects of this presumption are questionable.

First, the normal distribution does not seem to be the most adequate distribution to characterize the variables considered in this case (Y_i, X and ε_i). As Bernstein (1996) points out, normally distributed events are typical in natural phenomena but do not represent well facts derived from decisions made by people, such as in the field of economics and finance. Since Mandelbrot (1963) and Fama (1965), many empirical studies have corroborated this idea and have shown that, usually, returns of financial assets are not normally distributed. Other studies confirmed this for loan portfolios (e.g. Kalyvas et al, 2006 and Rosenberg and Schuermann, 2006). Bouyé et al. (2000) emphasize that, even though it is well known that asset returns are fat-tailed, people generally use normal processes to model financial returns because such methods have more tractable properties for computation.

Second, and likely the most important in the context of *portfolio* evaluations, expression (2.5) implicitly assumes normal dependence between Y_i and X (see Moreira, 2011) and therefore it is not able to identify different levels of connection among returns (or losses) in financial markets where extreme values tend to cluster. See, for instance, Embrechts et al. (2002) for financial assets in general and Di Clemente and Romano (2004) and Das and Geng (2006) for the specific case of credit portfolios.

Thus the assumption of normality (especially concerning Y) may lead to misestimated *PDs* since, as said before, many empirical studies have demonstrated that asset returns are seldom normally distributed. Furthermore, the assumptions in terms of X and ε are made for convenience and they may depart from the normality.

Some models have been proposed to relax the assumption of normality in (2.5) so that the calculation of the probability of default can take into account the higher proportion of data in the tails of the distributions (which results in higher *PDs*) when compared to the normal distribution. Notwithstanding, as we show below, these models are limited to changes in the univariate distributions of the variables and do not try to improve the dependence structure.

2.3. Some credit models suggested for non-normal variables

Starting from (2.2), Hull and White (2004) relax the distributions[3] of Y_i, X and ε_i, such that they can, for example, present heavy tails (which tends to increase the joint occurrences of extreme realizations of the latent variables when compared to the joint normal distribution). Representing the distributions of those three variables respectively by F, G and H and following the same steps that derived (2.5) from (2.2), the expression to estimate the probability of default ($\Pr[Y_i < y_c]$) conditional on the status $X = x^*$ turns into:

$$\Pr[Y_i < y_c \mid X = x^*] = H\left(\frac{F^{-1}(PD) - \sqrt{\rho_{ij}}\,G^{-1}(A_X^*)}{\sqrt{1-\rho_{ij}}}\right) \qquad 2.6$$

where A_X^* is the area below the analyzed economic scenario x^* in the distribution of X. PD is the (historical) probability of default and ρ_{ij} is the linear correlation between returns of obligors' assets. Obviously, the expression above cannot be solved unless the shapes of the three distributions F, G and H are known.

---

[2] This approach was used in the second Basel Accord (Basel II) and was kept in Basel III.

[3] Provided that they are scaled with mean 0 and variance 1.

Hull and White (2004) employ this model[4] (with X and ε_i following the Student t distribution) to estimate the joint probability of default of obligors in credit derivatives (collateralized debt obligations, CDOs and credit default swap, CDS). Assuming we know the distribution F of the historical probability of default, we can estimate the probability of default when $X = x^*$ as:

$$\Pr[Y_i < y_c \mid X = x^*] = T_v\left(\frac{F^{-1}(PD) - \sqrt{\rho_{ij}}T_v^{-1}(A_X^*)}{\sqrt{1-\rho_{ij}}}\right) \quad 2.7$$

where T_v is the Student t distribution with v degrees of freedom.

Bluhm et al. (2002), Kang and Shahabuddin (2005) and Kostadinov (2005) have also suggested the Student t distribution for X and ε_i to characterize the existence of more events (than the normal distribution) in the tails of credit portfolios' distributions. They obtain similar expressions that keep the basic structure of (2.6) and (2.7). Chan-Lau (2010) argues that the same reasoning could be applied in the context of the calculation of regulatory capital in financial institutions.

3. Inaccuracy of the Dependence Measure Used in the Models for Non-Normal Variables

Before we show that the dependence measure (correlation coefficient ρ_{ij}) used in the models presented in Section 2.3 is not supported by their derivation from (2.2), it is necessary to review some basic concepts.

3.1. Copulas and conditional distributions

Copulas are functions that link univariate distributions to form joint distributions of the variables considered which, in turn, give the probability that all variables are simultaneously below some specific values regardless of the shape of the univariate distributions:

$$\Pr[Y_1 < y_1,\dots,Y_n < y_n] = F_{1\dots n}(y_1,\dots,y_n) = C(F_1(y_1),\dots,F_n(y_n))$$

where *F(.)* represents a cumulative distribution function and C is a copula. Details on copulas can be found, for example, in Nelsen (2006) and Genest and Favre (2007).

The cumulative distribution of a random variable conditional on other variables is given by the first derivative of the copula that expresses the dependence among the variables with respect to the conditioning variables (see Joe, 1996, Aas et al., 2009 and Czado, 2010):

$$F(y \mid \mathbf{x}) = \frac{\partial C_{yx_j|\mathbf{x}_{-j}}(F(y \mid \mathbf{x}_{-j}), F(x_j \mid \mathbf{x}_{-j}))}{\partial F(x_j \mid \mathbf{x}_{-j})} \quad 3.1$$

where $F(y|\mathbf{x})$ is the distribution of Y evaluated at y and conditional on vector $\mathbf{x}$, $C_{yx_j|\mathbf{x}_{-j}}$ is a copula, $\mathbf{x_j}$ is a component of vector $\mathbf{x}$ and $\mathbf{x_{-j}}$ is the vector $\mathbf{x}$ excluding this component. When $\mathbf{x}$ is univariate, the conditional distribution is calculated as:

$$F(y \mid x) = C_{y|x}(F(y) \mid F(x)) = \frac{\partial C_{yx}(F(y), F(x))}{\partial F(x)} \quad 3.2$$

where y and x are the conditioned and the conditioning variables respectively and the remaining notation follows the preceding formula.

The Gaussian copula with normally-distributed marginals (which is implicit in some traditional credit risk models – see Li, 2000), for instance, has the first derivative given by (see Aas et al., 2009):

---

[4] In fact, these authors focus on the probability of the time of default, t_i, being smaller than a particular time t. As they state in their paper, this probability is equal to the probability of the obligors' asset returns Y_i being smaller than the level of the cutoff point y_c (which is the approach adopted here).

$$\Pr[Y < y \mid X = x] = F_{Y|X}(y \mid X = x\ ) = \Phi\left(\frac{\Phi^{-1}(F_Y(y)) - \theta_{YX}\Phi^{-1}(F_X(x))}{\sqrt{1-\theta_{YX}^2}}\right) \quad 3.3$$

where Φ and Φ^{-1} represent the standard normal distribution and its inverse respectively, *F(.|.)* is a conditional distribution, *F(.)* is an unconditional distribution and θ_{YX} is the Gaussian copula parameter[5] between *Y* and *X*.

3.2. The source of the shortcoming

As shown in Section 2, Eq. (2.5) is directly derived from (2.2). Recall that the coefficients ρ_{ij} and $\sqrt{1-\rho_{ij}}$ in (2.2) are valid only if all the three variables considered (Y_i, X and ε_i) follow the standard normal distribution. Hence, when this condition of normality is not met (even for only one of those three variables), we should start the derivation from (2.1) which implies the use of the unknown coefficients β_1 and β_2 rather than ρ_{ij} and $\sqrt{1-\rho_{ij}}$. Nonetheless, the models cited in Section 3 are derived from (2.2) even though the distributions of the variables are assumed to be different from the normal distribution (Student t in the examples mentioned). So, the use of the linear correlation in (2.6) and (2.7) is not supported by any statistical or mathematical property. Note that (2.3) still holds for non-normally distributed variables (scaled with zero mean and unit variance) but this has no effect on the calculation of the conditional probability of default since the derivation should start from (2.1).

Moreover, we can see that the probability of default conditional on $X = x^*$ expressed in (2.5), corresponds to the first derivative of the Gaussian copula (given by (3.3)) that connects Y_i to X where $\theta_{12} = \sqrt{\rho_{ij}}$. When the idiosyncratic risk (ε_i) is assumed to be normally distributed (function *H* in (2.6)), the dependence structure between Y_i and X is kept the same as in the traditional models for normal variables (Gaussian copula) even if those two variables (Y_i and X) are not normally distributed but, in this situation, θ_{12} is not necessarily equal to $\sqrt{\rho_{ij}}$ (because the conditional probability of default, $\Pr[Y_i < y_c \mid X = x^*]$, would be estimated from (2.1) and not from (2.2)).

When the idiosyncratic risk departs from the normality, as in (2.7) where ε_i follows the Student t distribution, the conditional probability of default is not associated with any copula family and, again, θ_{12} is not necessarily equal to $\sqrt{\rho_{ij}}$ (for the same reason stated above).

3.3. Impact of the inaccuracy on the estimation of credit losses in extreme scenarios

3.3.1. For a given dependence between the systematic factor and the latent variables

We ran simulations to test the impact of the theoretically inaccurate dependence measure $\sqrt{\rho_{ij}}$ on estimates of credit losses in an area where factor models are used in practice.

As mentioned in Section 2.1, expression (2.5) is suggested in Basel II to estimate the probability of default in exceptionally severe conditions (i.e. unexpected credit losses with confidence 0.999). However the assumption of normality for the variables considered limits the number of events in the tails of their distributions and, consequently, (2.5) tends to result in lower default probability than the losses observed in credit portfolios inasmuch as the literature has reported fat tails in loan returns (see Section 2.2).

[5] The parameter of the Gaussian copula is usually represented by ρ. We adopt the notation θ to distinguish the Gaussian copula parameter from the linear correlation coefficient between the variables studied. These two measures of dependence are identical only when the marginal distributions are normal.

According to the approach described in Section 2.3, a way to avoid the underestimation of the capital required to cover unexpected losses is to presume that the losses and the economic factor follow the Student t distribution (which has "fat tails"). In our simulations ahead we assume that the latent variables (Y) and the systematic factor (X) are t distributed with v degrees of freedom while, for convenience, the idiosyncratic risk (ε) is assumed to be normally distributed[6]. Thus, for a confidence level A_X^*, (2.7) can be written as:

$$[PD \mid X = x^*] = \Phi\left(\frac{T_v^{-1}(PD) + \sqrt{\rho_{ij}}\,T_v^{-1}(A_X^*)}{\sqrt{1-\rho_{ij}}}\right) \quad 3.4$$

where ρ_{ij} is the linear correlation between the latent variables that drive defaults in loans i and j, Φ represents the cumulative distribution function of the standard normal distribution and T_v^{-1} is the inverse of the standard Student t distribution with v degrees of freedom such that $T_v^{-1}(0.999) = -T_v^{-1}(0.001)$.

We tested five *PDs* (historical averages) 0.01, 0.03, 0.05, 0.07 and 0.10 (these are the expected losses in "typical" economic periods). For each *PD*, we considered four confidence levels (0.90, 0.95, 0.99 and 0.999 where the last one is the parameter specified in Basel) and three degrees of freedom (v) for the standard Student t distributions of X and Y (10, 20 and 30). The simulations of all scenarios were repeated 1,000 times to reduce potential randomness effects on our conclusions and the results presented ahead are the mean of each of the variables computed.

Since ε is assumed to be normally distributed, the dependence structure between the systematic factor X and each latent variable Y (i.e., Y_i and Y_j) that drives default is given by the Gaussian copula (with t-distributed margins in this case) as we can see by comparing (3.3) to (3.4). Thus we simulated pairs of standard Student t variables in a Gaussian dependence structure such that the "true" (underlying) dependence θ_{YX} between X and each Y resulted in specific values we wanted to test. Bear in mind that θ_{YX} is equivalent to the parameter θ_{YX} in (3.3) albeit, in that expression, the margins are normally distributed while in our simulations they follow Student t distributions. Hence, as said in Section 3.2, $\theta_{XY} = \sqrt{\rho_{ij}}$ when the margins are normal but this equality does not hold when the margins have different distributions (which is the case in our simulations).

Table 1 displays the probability of extreme credit losses estimated when θ_{YX} = 0.10. Each panel refers to a *PD* value. So, as an example, in Panel A, we have the simulations for *PD* = 0.01. The first column ("Conf") contains the four confidence levels. In the next six columns, each pair of columns is related to one of the three degrees of freedom of the standard Student t distributions of Y and X[7]. The second and third columns, for instance, give the results for v = 10. In this case, for the given (unobserved) θ_{YX} = 0.10, we found ρ_{ij} = 0.1576. Therefore, $\sqrt{\rho_{ij}}$ = 0.3770.

The column labeled "True" gives the unexpected losses computed when we plugged the "true" dependence (θ_{YX} = 0.10) in place of $\sqrt{\rho_{ij}}$ in (3.4) and corresponds to the default rate in the simulated data (a proxy for credit portfolios of financial institutions). The column "Estimated" displays the unexpected losses computed when we used $\sqrt{\rho_{ij}}$ in (3.4) as advocated by the models for non-normal variables presented in Section 2.3.

The unexpected losses predicted according to the latter approach were higher than the actual losses

[6] As can be inferred from (3.3), the assumption of normally-distributed ε leads to the first derivative of the bivariate Gaussian Copula (with two t-distributed margins) and this is essential for the simulations presented here.

[7] For simplicity, in each scenario, both distributions were assumed to have the same degrees of freedom but this presumption can be easily relaxed.

in the portfolios simulated for all confidence levels (i.e. the values in column "Estimated" were always greater than those in column "True") and this difference increased with the confidence level. So, the conservative parameter suggested in Basel (0.999) leads to the highest overestimation of credit losses. This behavior was observed in all other scenarios simulated. The probability of default estimated based on (3.4) was, on average, five times higher than the observed default rates but reached an estimate more than 13 times higher than the observed losses in one particular scenario (PD = 0.01, v = 10 and confidence level = 0.999). Generally speaking, the overestimations reached the highest degree when PD = 0.01 and decreased monotonically until PD = 0.10.

Table 1
Unexpected credit losses estimated by means of inaccurate dependence measures (Student t variables). "True" dependence θ_{YX} = 0.10

Panel A : *PD* = 0.01						
	v = 10		v = 20		v = 30	
	ρ_{ij} = 0.1576		ρ_{ij} = 0.1442		ρ_{ij} = 0.1463	
	θ_{YX} = 0.10	$\sqrt{\rho_{ij}}$ = 0.3770	θ_{YX} = 0.10	$\sqrt{\rho_{ij}}$ = 0.3517	θ_{YX} = 0.10	$\sqrt{\rho_{ij}}$ = 0.3545
Conf	"True"	Estimated	"True"	Estimated	"True"	Estimated
0.900	0.0041	0.0072	0.0041	0.0066	0.0041	0.0065
0.950	0.0047	0.0123	0.0046	0.0104	0.0046	0.0102
0.990	0.0062	0.0359	0.0058	0.0259	0.0057	0.0243
0.999	0.0091	0.1251	0.0078	0.0716	0.0074	0.0629
Panel B : PD = 0.03						
	v = 10		v = 20		v = 30	
	ρ_{ij} = 0.1357		ρ_{ij} = 0.1384		ρ_{ij} = 0.1400	
	θ_{YX} = 0.10	$\sqrt{\rho_{ij}}$ = 0.3404	θ_{YX} = 0.10	$\sqrt{\rho_{ij}}$ = 0.3406	θ_{YX} = 0.10	$\sqrt{\rho_{ij}}$ = 0.3415
Conf	"True"	Estimated	"True"	Estimated	"True"	Estimated
0.900	0.0231	0.0388	0.0229	0.0371	0.0228	0.0365
0.950	0.0256	0.0568	0.0251	0.0528	0.0249	0.0517
0.990	0.0319	0.1193	0.0303	0.1023	0.0297	0.0979
0.999	0.0431	0.2661	0.0381	0.2012	0.0365	0.1854
Panel C : PD = 0.05						
	v = 10		v = 20		v = 30	
	ρ_{ij} = 0.1252		ρ_{ij} = 0.1202		ρ_{ij} = 0.1216	
	θ_{YX} = 0.10	$\sqrt{\rho_{ij}}$ = 0.3248	θ_{YX} = 0.10	$\sqrt{\rho_{ij}}$ = 0.3150	θ_{YX} = 0.10	$\sqrt{\rho_{ij}}$ = 0.3170
Conf	"True"	Estimated	"True"	Estimated	"True"	Estimated
0.900	0.0461	0.0743	0.0457	0.0706	0.0455	0.0701
0.950	0.0505	0.1016	0.0497	0.0936	0.0494	0.0923
0.990	0.0613	0.1844	0.0586	0.1567	0.0577	0.1514
0.999	0.0799	0.3439	0.0716	0.2640	0.0693	0.2474
Panel D : PD = 0.07						
	v = 10		v = 20		v = 30	
	ρ_{ij} = 0.1291		ρ_{ij} = 0.1210		ρ_{ij} = 0.1202	
	θ_{YX} = 0.10	$\sqrt{\rho_{ij}}$ = 0.3353	θ_{YX} = 0.10	$\sqrt{\rho_{ij}}$ = 0.3226	θ_{YX} = 0.10	$\sqrt{\rho_{ij}}$ = 0.3230
Conf	"True"	Estimated	"True"	Estimated	"True"	Estimated
0.900	0.0704	0.1132	0.0698	0.1076	0.0695	0.1065
0.950	0.0765	0.1499	0.0753	0.1379	0.0749	0.1356
0.990	0.0913	0.2524	0.0875	0.2155	0.0863	0.2075
0.999	0.1162	0.4292	0.1051	0.3375	0.1020	0.3157

Table 1 (continued)
Unexpected credit losses estimated by means of inaccurate dependence measures (Student t variables). "True" dependence θ_{YX} = 0.10

Panel E : PD = 0.10						
	v = 10		v = 20		v = 30	
	ρ_{ij} = 0.1321		ρ_{ij} = 0.1192		ρ_{ij} = 0.1247	
	θ_{YX} = 0.10	$\sqrt{\rho_{ij}}$ = 0.3321	θ_{YX} = 0.10	$\sqrt{\rho_{ij}}$ = 0.3137	θ_{YX} = 0.10	$\sqrt{\rho_{ij}}$ = 0.3270
Conf	"True"	Estimated	"True"	Estimated	"True"	Estimated
0.900	0.1073	0.1678	0.1065	0.1584	0.1062	0.1602
0.950	0.1157	0.2148	0.1140	0.1965	0.1135	0.1984
0.990	0.1355	0.3342	0.1304	0.2869	0.1288	0.2868
0.999	0.1681	0.5087	0.1536	0.4136	0.1495	0.4067

Notes: *PD* is the (historical) average probability of default. θ_{YX} is the "true" dependence between the latent variable (Y_i in expression (2.7) which is implicit in (3.4)) that drives default and the systematic factor (X in (2.7) and (3.4)). The dependence structure between X and each Y is the Gaussian copula such that the calculation of the extreme losses can be done by simply changing the marginals of Y and X in (2.5). ρ_{ij} is the linear correlation between latent variables Y_i and Y_j. $\sqrt{\rho_{ij}}$ is the dependence measure used to express the underlying "true" dependence θ_{YX}. The degrees of freedom of the Student t distributions simulated are represented by v. "True" and estimated *PDs* are the probabilities of default estimated according to (3.4) for extreme scenarios (confidence levels "Conf") by means of the "true" dependence parameter (θ_{YX}) and the approximation $\sqrt{\rho_{ij}}$, respectively.

We also tested other "true" dependence levels between X and Y and found analogous results. Table 2 presents the estimates for θ_{YX} = 0.25 and the results for other values of θ_{YX} are available upon request.

Table 2
Unexpected credit losses estimated by means of inaccurate dependence measures (Student t variables). "True" dependence θ_{YX} = 0.25

Panel A : PD = 0.01						
	v = 10		v = 20		v = 30	
	ρ_{ij} = 0.1501		ρ_{ij} = 0.1620		ρ_{ij} = 0.1601	
	θ_{YX} = 0.25	$\sqrt{\rho_{ij}}$ = 0.3606	θ_{YX} = 0.25	$\sqrt{\rho_{ij}}$ = 0.3710	θ_{YX} = 0.25	$\sqrt{\rho_{ij}}$ = 0.3733
Conf	"True"	Estimated	"True"	Estimated	"True"	Estimated
0.900	0.0062	0.0070	0.0060	0.0066	0.0059	0.0066
0.950	0.0085	0.0118	0.0080	0.0108	0.0078	0.0106
0.990	0.0162	0.0348	0.0138	0.0287	0.0132	0.0262
0.999	0.0372	0.1226	0.0264	0.0832	0.0239	0.0694
Panel B : PD = 0.03						
	v = 10		v = 20		v = 30	
	ρ_{ij} = 0.1450		ρ_{ij} = 0.1446		ρ_{ij} = 0.1393	
	θ_{YX} = 0.25	$\sqrt{\rho_{ij}}$ = 0.3578	θ_{YX} = 0.25	$\sqrt{\rho_{ij}}$ = 0.3581	θ_{YX} = 0.25	$\sqrt{\rho_{ij}}$ = 0.3498
Conf	"True"	Estimated	"True"	Estimated	"True"	Estimated
0.900	0.0332	0.0400	0.0323	0.0382	0.0320	0.0373
0.950	0.0425	0.0591	0.0405	0.0547	0.0399	0.0524
0.990	0.0699	0.1254	0.0621	0.1057	0.0599	0.0974
0.999	0.1313	0.2825	0.1014	0.2080	0.0940	0.1825

Table 2 (continued)

Unexpected credit losses estimated by means of inaccurate dependence measures (Student t variables). "True" dependence θ_{YX} = 0.25

Panel C : PD = 0.05						
	v = 10		v = 20		v = 30	
	ρ_{ij} = 0.1608		ρ_{ij} = 0.1680		ρ_{ij} = 0.1703	
	θ_{YX} = 0.25	$\sqrt{\rho_{ij}}$ = 0.3801	θ_{YX} = 0.25	$\sqrt{\rho_{ij}}$ = 0.3841	θ_{YX} = 0.25	$\sqrt{\rho_{ij}}$ = 0.3872
Conf	"True"	Estimated	"True"	Estimated	"True"	Estimated
0.900	0.0645	0.0813	0.0630	0.0786	0.0626	0.0779
0.950	0.0801	0.1154	0.0768	0.1095	0.0758	0.1078
0.990	0.1232	0.2207	0.1114	0.1963	0.1080	0.1893
0.999	0.2109	0.4230	0.1698	0.3442	0.1592	0.3222
Panel D : PD = 0.07						
	v = 10		v = 20		v = 30	
	ρ_{ij} = 0.1722		ρ_{ij} = 0.1750		ρ_{ij} = 0.1734	
	θ_{YX} = 0.25	$\sqrt{\rho_{ij}}$ = 0.3847	θ_{YX} = 0.25	$\sqrt{\rho_{ij}}$ = 0.3962	θ_{YX} = 0.25	$\sqrt{\rho_{ij}}$ = 0.3907
Conf	"True"	Estimated	"True"	Estimated	"True"	Estimated
0.900	0.0966	0.1228	0.0946	0.1203	0.0939	0.1180
0.950	0.1175	0.1704	0.1132	0.1625	0.1117	0.1581
0.990	0.1731	0.3037	0.1582	0.2718	0.1536	0.2590
0.999	0.2791	0.5100	0.2305	0.4347	0.2173	0.4045
Panel E : PD = 0.10						
	v = 10		v = 20		v = 30	
	ρ_{ij} = 0.1513		ρ_{ij} = 0.1383		ρ_{ij} = 0.1539	
	θ_{YX} = 0.25	$\sqrt{\rho_{ij}}$ = 0.3655	θ_{YX} = 0.25	$\sqrt{\rho_{ij}}$ = 0.3452	θ_{YX} = 0.25	$\sqrt{\rho_{ij}}$ = 0.3675
Conf	"True"	Estimated	"True"	Estimated	"True"	Estimated
0.900	0.1439	0.1766	0.1413	0.1663	0.1404	0.1704
0.950	0.1712	0.2295	0.1657	0.2098	0.1639	0.2161
0.990	0.2407	0.3642	0.2225	0.3125	0.2172	0.3216
0.999	0.3640	0.5607	0.3089	0.4530	0.2940	0.4598

Notes: *PD* is the (historical) average probability of default. θ_{YX} is the "true" dependence between the latent variable (Y_i in expression (2.7) which is implicit in (3.4)) that drives default and the systematic factor (X in (2.7) and (3.4)). The dependence structure between X and each Y is the Gaussian copula such that the calculation of the extreme losses can be done by simply changing the marginals of Y and X in (2.5). ρ_{ij} is the linear correlation between latent variables Y_i and Y_j. $\sqrt{\rho_{ij}}$ is the dependence measure used to express the underlying "true" dependence θ_{YX}. The degrees of freedom of the Student t distributions simulated are represented by v. "True" and estimated *PDs* are the probabilities of default estimated according to (3.4) for extreme scenarios (confidence levels) by means of the "true" dependence parameter (θ_{YX}) and the approximation $\sqrt{\rho_{ij}}$, respectively.

In principle, overestimated credit losses could be thought as advantageous in this regulatory environment but when they are excessive, as it is in some scenarios of our simulations, they become a problem for financial institutions given that unnecessary capital held as a buffer against potential losses lessens the amount of resources available to investments and resulting profits.

As a robustness check, we ran additional simulations based on the dependence structure implied in (2.5) to confirm whether the dependence measure $\sqrt{\rho_{ij}}$ is really equal to θ_{YX} when all the margins are normally distributed. The results are reported in Table 3 for the same parameters considered above (θ_{YX}, *PD*, v and confidence). Panels A and B are respectively related to θ_{YX} = 0.10

Table 3
Unexpected credit losses estimated when all variables are normally distributed (for given θ_{YX})

Panel A: θ_{YX} = 0.10

	PD = 0.01		PD = 0.03		PD = 0.05		PD = 0.07		PD = 0.10	
ρ_{ij} = 0.01	θ_{YX} = 0.10	$\sqrt{\rho_{ij}}$ = 0.10	θ_{YX} = 0.10	$\sqrt{\rho_{ij}}$ = 0.10	θ_{YX} = 0.10	$\sqrt{\rho_{ij}}$ = 0.10	θ_{YX} = 0.10	$\sqrt{\rho_{ij}}$ = 0.10	θ_{YX} = 0.10	$\sqrt{\rho_{ij}}$ = 0.10
Confidence	"True"	Estimated	"True"	Estimated	"True"	Estimated	"True"	Estimated	"True"	Estimated
0.900	0.0136	0.0136	0.0391	0.0391	0.0637	0.0637	0.0877	0.0877	0.1231	0.1231
0.950	0.0149	0.0149	0.0423	0.0423	0.0684	0.0684	0.0937	0.0937	0.1307	0.1307
0.990	0.0176	0.0176	0.0489	0.0489	0.0778	0.0778	0.1056	0.1056	0.1458	0.1458
0.999	0.0212	0.0212	0.0572	0.0572	0.0896	0.0896	0.1202	0.1202	0.1640	0.1640

Panel B: θ_{YX} = 0.25

	PD = 0.01		PD = 0.03		PD = 0.05		PD = 0.07		PD = 0.10	
ρ_{ij} = 0.0625	θ_{YX} = 0.25	$\sqrt{\rho_{ij}}$ = 0.25	θ_{YX} = 0.25	$\sqrt{\rho_{ij}}$ = 0.25	θ_{YX} = 0.25	$\sqrt{\rho_{ij}}$ = 0.25	θ_{YX} = 0.25	$\sqrt{\rho_{ij}}$ = 0.25	θ_{YX} = 0.25	$\sqrt{\rho_{ij}}$ = 0.25
Confidence	"True"	Estimated	"True"	Estimated	"True"	Estimated	"True"	Estimated	"True"	Estimated
0.900	0.0192	0.0192	0.0535	0.0535	0.0857	0.0857	0.1164	0.1164	0.1604	0.1604
0.950	0.0240	0.0240	0.0645	0.0645	0.1013	0.1013	0.1358	0.1358	0.1844	0.1844
0.990	0.0358	0.0358	0.0898	0.0898	0.1361	0.1361	0.1780	0.1780	0.2349	0.2349
0.999	0.0544	0.0544	0.1261	0.1261	0.1839	0.1839	0.2341	0.2341	0.2996	0.2996

Notes: PD is the (historical) average probability of default. θ_{YX} is the "true" dependence between the latent variable (Y_i in expression (2.7) which is implicit in (3.4)) that drives default and the systematic factor (*X* in (2.7) and (3.4)). The dependence structure between *X* and each *Y* is the Gaussian copula such that the calculation of the extreme losses can be done via (2.5). ρ_{ij} is the linear correlation between latent variables Y_i and Y_j. $\sqrt{\rho_{ij}}$ is the dependence measure used to express the underlying "true" dependence θ_{YX}. "True" and estimated *PDs* are the probabilities of default estimated according to (3.4) for extreme scenarios (confidence levels) by means of the "true" dependence parameter (θ_{YX}) and the approximation $\sqrt{\rho_{ij}}$, respectively. In this table, the results are the same for the "true" and the estimated losses since all the variables are normal and, consequently, $\sqrt{\rho_{ij}} = \theta_{YX}$ (up to the fourth decimal place in our simulations).

and θ_{YX} = 0.25. We see that $\sqrt{\rho_{ij}} = \theta_{YX}$ in all scenarios and, therefore, the losses estimated according to (2.5) are the same for both dependence measures[8].

3.3.2. For a given correlation among latent variables

In the previous section, we simulated the variables X, Y_i and Y_j as if we knew the dependence between them (θ_{YX}). After that, we checked the resultant correlation ρ_{ij} between Y_i and Y_j and then compared potential losses estimated via $\sqrt{\rho_{ij}}$ with losses estimated by means of θ_{YX}.

Nonetheless a more realistic approach would be to start from a given correlation between the latent variables (Y_i and Y_j) since θ_{YX} is not observable. The value of ρ_{ij} for different credit classes was calibrated in Basel II. For credit cards and mortgages, e.g., ρ_{ij} is equal to 0.04 and 0.15, respectively. The correlations in other classes are estimated as a function of *PD* (see BCBS, 2006 for more details).

In this section, we simulated standard t-distributed variables X, Y_i and Y_j such that ρ_{ij} resulted in some specific values given in Basel (the aforementioned correlations for credit cards and mortgages). Then we checked the resultant θ_{YX} and used it to compute the "true" extreme losses which were compared to the losses estimated in accordance with the assumption that $\sqrt{\rho_{ij}}$ represents relationship between X and each Y.

As in Section 3.3.1, we analyzed five (historical) average *PDs* (0.01, 0.03, 0.05, 0.07 and 0.10), three values for the degree of freedom of the Student t distributions of X, Y_i and Y_j (10, 20 and 30) and four confidence levels (0.90, 0.95, 0.99 and 0.999). See Tables 4 and 5 for ρ_{ij} = 0.04 and ρ_{ij} = 0.15, respectively, where each panel pertains to a *PD* value.

The results corroborated our prior findings since, for all scenarios, the losses estimated by means of the dependence parameter $\sqrt{\rho_{ij}}$ were larger than the ('true") losses observed in the simulated data (i.e., $\sqrt{\rho_{ij}} > \theta_{YX}$). The main difference was that the overestimation level was roughly constant for all *PDs* tested (whilst such level decreased with *PD* in Tables 1 and 2). Therefore, this is additional evidence in favor of our conclusion that the models cited in Section 2.3 tend to overestimate extreme losses. It is interesting to note that θ_{YX} decreased monotonically with *PD* levels.

Further simulations for normally distributed variables (as in Table 3) confirmed that, when we start from a given linear correlation between the latent variables (ρ_{ij}), the resultant θ_{YX} equals to $\sqrt{\rho_{ij}}$ (up to the fourth decimal place in our simulations). For the sake of brevity, the unexpected losses estimated in these conditions are not presented here.

Table 4
Unexpected credit losses estimated by means of inaccurate dependence measures (Student t variables). "True" correlation ρ_{ij} = 0.04

Panel A : PD = 0.01						
	v = 10		v = 20		v = 30	
	ρ_{ij} = 0.04		ρ_{ij} = 0.04		ρ_{ij} = 0.04	
	θ_{YX} = 0.1722	$\sqrt{\rho_{ij}}$ = 0.20	θ_{YX} = 0.1722	$\sqrt{\rho_{ij}}$ = 0.20	θ_{YX} = 0.1722	$\sqrt{\rho_{ij}}$ = 0.20
Conf	"True"	Estimated	"True"	Estimated	"True"	Estimated
0.900	0.0051	0.0056	0.0050	0.0055	0.0044	0.0053
0.950	0.0064	0.0072	0.0061	0.0070	0.0051	0.0066
0.990	0.0101	0.0123	0.0090	0.0112	0.0068	0.0101
0.999	0.0187	0.0248	0.0145	0.0195	0.0094	0.0163

[8] This equality was observed up to the fourth decimal place.

Table 4 (continued)

Unexpected credit losses estimated by means of inaccurate dependence measures (Student t variables). "True" correlation ρ_{ij} = 0.04

Panel B : PD = 0.03						
	v = 10		v = 20		v = 30	
	ρ_{ij} = 0.04		ρ_{ij} = 0.04		ρ_{ij} = 0.04	
	θ_{YX} = 0.1457	$\sqrt{\rho_{ij}}$ = 0.20	θ_{YX} = 0.1457	$\sqrt{\rho_{ij}}$ = 0.20	θ_{YX} = 0.1457	$\sqrt{\rho_{ij}}$ = 0.20
Conf	"True"	Estimated	"True"	Estimated	"True"	Estimated
0.900	0.0261	0.0294	0.0257	0.0281	0.0256	0.0277
0.950	0.0303	0.0357	0.0294	0.0332	0.0292	0.0324
0.990	0.0413	0.0531	0.0383	0.0459	0.0374	0.0439
0.999	0.0626	0.0896	0.0526	0.0673	0.0500	0.0619
Panel C : PD = 0.05						
	v = 10		v = 20		v = 30	
	ρ_{ij} = 0.04		ρ_{ij} = 0.04		ρ_{ij} = 0.04	
	θ_{YX} = 0.1353	$\sqrt{\rho_{ij}}$ = 0.20	θ_{YX} = 0.1353	$\sqrt{\rho_{ij}}$ = 0.20	θ_{YX} = 0.1353	$\sqrt{\rho_{ij}}$ = 0.20
Conf	"True"	Estimated	"True"	Estimated	"True"	Estimated
0.900	0.0503	0.0595	0.0497	0.0589	0.0494	0.0586
0.950	0.0569	0.0715	0.0555	0.0699	0.0551	0.0693
0.990	0.0733	0.1037	0.0689	0.0968	0.0676	0.0946
0.999	0.1032	0.1674	0.0895	0.1412	0.0858	0.1340
Panel D : PD = 0.07						
	v = 10		v = 20		v = 30	
	ρ_{ij} = 0.04		ρ_{ij} = 0.04		ρ_{ij} = 0.04	
	θ_{YX} = 0.1219	$\sqrt{\rho_{ij}}$ = 0.20	θ_{YX} = 0.1219	$\sqrt{\rho_{ij}}$ = 0.20	θ_{YX} = 0.1219	$\sqrt{\rho_{ij}}$ = 0.20
Conf	"True"	Estimated	"True"	Estimated	"True"	Estimated
0.900	0.0740	0.0831	0.0732	0.0839	0.0729	0.0841
0.950	0.0819	0.0958	0.0803	0.0963	0.0798	0.0964
0.990	0.1010	0.1278	0.0960	0.1252	0.0945	0.1241
0.999	0.1343	0.1862	0.1192	0.1700	0.1152	0.1651
Panel E : PD = 0.10						
	v = 10		v = 20		v = 30	
	ρ_{ij} = 0.04		ρ_{ij} = 0.04		ρ_{ij} = 0.04	
	θ_{YX} =	$\sqrt{\rho_{ij}}$ = 0.20	θ_{YX} =	$\sqrt{\rho_{ij}}$ = 0.20	θ_{YX} =	$\sqrt{\rho_{ij}}$ = 0.20
Conf	"True"	Estimated	"True"	Estimated	"True"	Estimated
0.900	0.1074	0.1295	0.1065	0.1310	0.1063	0.1314
0.950	0.1159	0.1487	0.1141	0.1499	0.1136	0.1501
0.990	0.1357	0.1963	0.1306	0.1931	0.1291	0.1918
0.999	0.1685	0.2798	0.1539	0.2580	0.1499	0.2514

Notes: PD is the (historical) average probability of default. θ_{YX} is the "true" dependence between the latent variable (Y_i in expression (2.7) which is implicit in (3.4)) that drives default and the systematic factor (X in (2.7) and (3.4)). The dependence structure between X and each Y is the Gaussian copula such that the calculation of the extreme losses can be done by simply changing the marginals of Y and X in (2.5). ρ_{ij} is the linear correlation between latent variables Y_i and Y_j (in this case, it is set equal to 0.04, which is the value defined in Basel Accord for credit cards). $\sqrt{\rho_{ij}}$ is the dependence measure used to express the underlying "true" dependence θ_{YX}. The degrees of freedom of the Student t distributions simulated are represented by v. "True" and estimated PDs are the probabilities of default estimated according to (3.4) for extreme scenarios (confidence levels) by means of the "true" dependence parameter (θ_{YX}) and the approximation $\sqrt{\rho_{ij}}$, respectively.

Table 5

Unexpected credit losses estimated by means of inaccurate dependence measures (Student t variables). "True" correlation ρ_{ij} = 0.15

Panel A : PD = 0.01						
	v = 10		v = 20		v = 30	
	ρ_{ij} = 0.15		ρ_{ij} = 0.15		ρ_{ij} = 0.15	
	θ_{YX} = 0.2916	$\sqrt{\rho_{ij}}$ = 0.3873	θ_{YX} = 0.2916	$\sqrt{\rho_{ij}}$ = 0.3873	θ_{YX} = 0.2916	$\sqrt{\rho_{ij}}$ = 0.3873
Conf	"True"	Estimated	"True"	Estimated	"True"	Estimated
0.90	0.0067	0.0078	0.0065	0.0073	0.0064	0.0071
0.95	0.0097	0.0129	0.0091	0.0114	0.0088	0.0110
0.99	0.0203	0.0345	0.0171	0.0260	0.0162	0.0238
0.999	0.0520	0.1105	0.0354	0.0644	0.0316	0.0548
Panel B : PD = 0.03						
	v = 10		v = 20		v = 30	
	ρ_{ij} = 0.15		ρ_{ij} = 0.15		ρ_{ij} = 0.15	
	θ_{YX} = 0.2221	$\sqrt{\rho_{ij}}$ = 0.3873	θ_{YX} = 0.2221	$\sqrt{\rho_{ij}}$ = 0.3873	θ_{YX} = 0.2221	$\sqrt{\rho_{ij}}$ = 0.3873
Conf	"True"	Estimated	"True"	Estimated	"True"	Estimated
0.90	0.0313	0.0420	0.0306	0.0406	0.0303	0.0402
0.95	0.0391	0.0609	0.0374	0.0575	0.0369	0.0564
0.99	0.0612	0.1237	0.0549	0.1076	0.0532	0.1029
0.999	0.1092	0.2778	0.0861	0.2090	0.0803	0.1911
Panel C : PD = 0.05						
	v = 10		v = 20		v = 30	
	ρ_{ij} = 0.15		ρ_{ij} = 0.15		ρ_{ij} = 0.15	
	θ_{YX} = 0.2002	$\sqrt{\rho_{ij}}$ = 0.3873	θ_{YX} = 0.2002	$\sqrt{\rho_{ij}}$ = 0.3873	θ_{YX} = 0.2002	$\sqrt{\rho_{ij}}$ = 0.3873
Conf	"True"	Estimated	"True"	Estimated	"True"	Estimated
0.90	0.0583	0.0817	0.0572	0.0795	0.0568	0.0788
0.95	0.0695	0.1129	0.0671	0.1075	0.0664	0.1058
0.99	0.0993	0.2069	0.0912	0.1837	0.0888	0.1768
0.999	0.1578	0.4030	0.1304	0.3193	0.1234	0.2965
Panel D : PD = 0.07						
	v = 10		v = 20		v = 30	
	ρ_{ij} = 0.15		ρ_{ij} = 0.15		ρ_{ij} = 0.15	
	θ_{YX} = 0.1987	$\sqrt{\rho_{ij}}$ = 0.3873	θ_{YX} = 0.1987	$\sqrt{\rho_{ij}}$ = 0.3873	θ_{YX} = 0.1987	$\sqrt{\rho_{ij}}$ = 0.3873
Conf	"True"	Estimated	"True"	Estimated	"True"	Estimated
0.90	0.0873	0.1221	0.0858	0.1194	0.0853	0.1185
0.95	0.1024	0.1634	0.0992	0.1568	0.0983	0.1547
0.99	0.1411	0.2797	0.1307	0.2529	0.1277	0.2449
0.999	0.2132	0.4968	0.1800	0.4094	0.1712	0.3849
Panel E : PD = 0.10						
	v = 10		v = 20		v = 30	
	ρ_{ij} = 0.15		ρ_{ij} = 0.15		ρ_{ij} = 0.15	
	θ_{YX} = 0.1758	$\sqrt{\rho_{ij}}$ = 0.3873	θ_{YX} = 0.1758	$\sqrt{\rho_{ij}}$ = 0.3873	θ_{YX} = 0.1758	$\sqrt{\rho_{ij}}$ = 0.3873
Conf	"True"	Estimated	"True"	Estimated	"True"	Estimated
0.90	0.1253	0.1797	0.1236	0.1756	0.1231	0.1743
0.95	0.1423	0.2315	0.1388	0.2222	0.1377	0.2193
0.99	0.1840	0.3671	0.1730	0.3343	0.1698	0.3246
0.999	0.2566	0.5920	0.2238	0.5005	0.2150	0.4742

Notes: *PD* is the (historical) average probability of default. θ_{YX} is the "true" dependence between the latent variable (Y_i in expression (2.7) which is implicit in (3.4)) that drives default and the systematic factor (X in (2.7) and (3.4)). The dependence structure between X and each Y is the Gaussian copula such that the calculation of the extreme losses can be done by simply changing the marginals of Y and X in (2.5). ρ_{ij} is the linear correlation between latent variables Y_i and Y_j (in this case, it is set equal to 0.15, which is the value defined in Basel Accord for mortgages). $\sqrt{\rho_{ij}}$ is the dependence measure used to express the underlying "true" dependence θ_{YX}. The degrees of freedom of the Student t distributions simulated are represented by v. "True" and estimated *PDs* are the probabilities of default estimated according to (3.4) for extreme scenarios (confidence levels) by means of the "true" dependence parameter (θ_{YX}) and the approximation $\sqrt{\rho_{ij}}$, respectively.

4. Suggestions Towards a Solution

Given that the change of the distributions of Y_i, X and ε_i and the estimation of the conditional probability of default derived from (2.1) or (2.2) are not compatible with the Gaussian copula, one way out could be the use of the first derivative of other copula families (following (3.2)[9]) which may capture distinct relationship structures, such as tail dependence, between Y_i and X irrespective of their distributions. The first derivative of some bivariate copulas are presented, e.g., in Joe (1997, Chapter 5) and in Aas et al. (2009, Appendix C). In the case of the Gaussian copula with normal margins (as in (3.3)), the first derivative has the form:

$$\Pr[Y_i < y_c \mid X = x^*] = \Phi\left(\frac{\Phi^{-1}(PD) - \theta_{YX}\Phi^{-1}(A_X^*)}{\sqrt{1-\theta_{YX}^2}}\right) = \Phi\left(\frac{\Phi^{-1}(PD) - \sqrt{\rho_{ij}}\Phi^{-1}(A_X^*)}{\sqrt{1-\rho_{ij}}}\right)$$

where θ_{YX} is the parameter of the copula related to Y_i and X. In this case, this parameter is equal to the linear correlation between those two variables ($\theta_{YX} = \rho_{YX}$) which, according to (2.3), is equal to the square root of the linear correlation between the latent variables Y of two equicorrelated loans i and j ($\sqrt{\rho_{ij}} = \rho_{YX}$). Thus, if we can calculate ρ_{ij} (which is assumed, for example, in Basel Accords), we can find the copula parameter θ_{YX} and then calculate the conditional probability of default. The remaining notation follows (3.3).

However, this relationship across the abovementioned dependence measures is not valid when we employ the Gaussian structure with non-normal margins (as in (2.7) when H is normally distributed and F and G have other distributions) or any other copula structure and, therefore, it is not possible to infer the copula parameter that indicates the dependence between Y_i and X from a dependence measure between loans' latent variables (Y_i and Y_j, for example). So, in these situations, we face the challenge of estimating the copula parameter between Y_i and X based on a dependence measure (linear correlation or rank correlation, for instance) across the latent variables of pairs of loans (Y_i and Y_j, for example). Up to this point, to the best of our knowledge, the link between the dependence measures pertaining to Y_i and X and to Y_i and Y_j is unknown apart from the case of the Gaussian structure with normally distributed margins. So, finding this link seems to be the next step towards a solution to define accurate dependence measures and to estimate the probability of default conditional on specific events or on specific economic scenarios without assuming that all variables are normally distributed.

5. Conclusions

Many popular credit risk models assume that returns of obligors' assets are normally distributed, not only individually (univariate normal distribution for each debtor's asset returns) but

[9] If multiple factors are assumed to impact the latent variable Y_i, (3.1) should be used.

also at the portfolio level (joint distribution of asset returns represented by the multivariate normal). However it is well known that asset returns (loans included) are not normally distributed and present tail dependence. Therefore these traditional approaches are not able to capture possible strong association among high losses and are prone to underestimate the probability of joint extreme losses.

Some models have relaxed the assumption of univariate normality and therefore have the advantage of identifying more occurrences of extreme values (when compared to methods founded on the normal distribution). Nonetheless we showed that, in spite of this benefit, such models are based on dependence measures incompatible with some presumptions implicit in the formula used in their derivation and, according to our simulations, this bias results in considerable overestimation of losses in some cases (especially for low default probabilities and high confidence levels).

As underestimated losses are a problem in risk management, excessively overestimated losses also have a downside in some circumstances. This is the case of the computation of the capital required to cover unexpected credit losses in financial institutions. Even though this might seem to be interesting from the regulatory standpoint, when institutions hold excessive capital (more than effectively necessary to cover losses) they miss opportunities of investing resources and profiting from them.

The models that incorporate non-normal variables are limited to changes in the marginal distributions and do not analyze the dependence structure (copula) between the systematic factor and the latent variables that drive defaults. An alternative way to relax both assumptions of normality (univariate and multivariate) and still to guarantee accurate dependence parameters is to use conditional distributions which are given by the first derivatives of copulas. However this solution is not complete yet as we do not know the connection between a dependence measure related to the loans (or their latent variables) and a dependence measure that associates the systematic (economic) factor to the latent variable of each loan. So, while this drawback persists, the models mentioned in Section 2.3 remain as an option to avoid the potential underestimation of the probability of default due to the unrealistic assumption of normality but users must keep in mind the limitation concerning the imprecision of the dependence measure and the consequent overestimation.

References

Aas, K., C. Czado, A. Frigessi and H. Bakken, 2009, Pair-copula constructions of multiple dependence, *Insurance Mathematics & Economics*, 44, 182-198.

Basel Committee on Banking Supervision - BCBS (2006). International Convergence of Capital Measurement and Capital Standards: A Revised Framework. Bank for International Settlements Report.

Bernstein, Peter L. (1996). *Against the Gods. The Remarkable Story of Risk*. New York: John Wiley & Sons.

Bluhm, C., L. Overbeck and C. Wagner, 2002, *An Introduction to Credit Risk Modeling*, London: Chapman & Hall/CRC.

Bouyé, E., V. Durrleman, A. Nikeghbali, G. Riboulet, T. Roncalli (2000). Copulas for Finance: A Reading Guide and Some Applications. *Working Paper*.

Chan-Lau, Jorge A., 2010, Regulatory Capital Charges for Too-Connected-to-Fail Institutions: A Practical Proposal, *Financial Markets, Institutions and Instruments*, 19, Issue 5, 355–379.

Crouhy, M., D. Galai and R. Mark, 2000, A comparative analysis of current credit risk models, *Journal of Banking & Finance*, 24, 59-117.

Czado, Claudia (2010). Pair-copula constructions for multivariate copulas. In: P. Jaworki, F. Durante, W. Härdle and T. Rychlik (eds.), *Workshop on Copula Theory and its Applications*, Dortrecht: Springer, 93-110.

Das, S., G. Geng (2006). Correlated Default Processes: A Criterion-Based Copula Approach. In: Fong, H. (ed.) "*The Credit Market Handbook. Advanced Modeling Issues*". Hoboken, New Jersey: John Wiley & Sons, pp. 186-218.

Di Clemente, A., C. Romano (2004). Measuring and Optimizing Portfolio Credit Risk: A Copula-based Approach. *Economic Notes (by Banca Monte dei Paschi di Siena SpA)*, Vol. 33, no 3, pp. 325-357.

Embrechts, P., A. McNeil, D. Strautman (2002). Correlation and Dependency in Risk Management: Properties and Pitfalls. In: Dempster, M.A.H. (ed.) "*Risk Management: Value at Risk and Beyond*". Cambridge: Cambridge University Press, pp. 176-223.

Fama, E. F., 1965, The behavior of stock-market prices, *Journal of Business*, 38(1), 34-105.

Genest, C., A-C. Favre (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering* - July/August, pp. 347-368.

Hull, J. and A. White, 2004, Valuation of a CDO and an n^{th} to Default CDS Without Monte Carlo Simulation, *Journal of Derivatives*, 12, 2, Winter.

Joe, Harry, 1996, Families of m-variate distributions with given margins and m(m-1)/2 bivariate dependence parameters, *IMS Lecture Notes*, 28, 120-141.

Joe, Harry, 1997, *Multivariate Models and Dependence Concepts*, London, New York: Chapman & Hall/CRC.

Kalyvas, L., I. Akkizdis, I. Zourka, V. Bouchereau (2006). Integrating Market, Credit and Operational Risk. A complete guide for bankers and risk professionals. Pamplona, Spain: Risk Books.

Kang, W. and P. Shahabuddin, 2005, Fast simulation for multifactor portfolio credit risk in the t-copula model, *Proceedings of the 2005 Winter Simulation Conference*, 1859-1868.

Kostadinov, Krassimir, 2005, Non-parametric Estimation of Elliptical Copulae With Application to Credit Risk, *Working paper*, Munich University Technology, http://www.ma.tum.de/stat.

Li, David (2000). On Default Correlation: A Copula Function Approach, *Journal of Fixed Income,* 9, 43-55.

Mandelbrot, B. B., 1963, The variation of certain speculative prices, *Journal of Business*, 36(4), 394-419.

Merton, Robert C., 1974, On the Pricing of Corporate Debt: The Risk Structure of Interest Rates, *Journal of Finance*, 28, 449-470.

Moreira, Fernando F., 2011, Enhancing Basel Method Via Conditional Distributions that Capture Stronger Connection Among Credit Losses in Downturns, *Working paper*, Available at SSRN: http://ssrn.com/abstract= 1784242.

Nelsen, Roger B. (2006). *An Introduction to Copulas*. New York: Springer. 2nd edition.

Owen, D. and G. Steck, 1962, Moments of Order Statistics from the Equicorrelated Multivariate Normal Distribution. *The Annals of Mathematical Statistics*, 33(4), 1286-1291.

Perli, R. and W. Nayda, 2004, Economic and regulatory capital allocation for revolving retail exposures, *Journal of Banking & Finance*, 28, 789-809.

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

Schönbucher, Philipp J., 2000, Factor Models for Portfolio Credit Risk, *Discussion Paper* 16/2001, Bonn Graduate School of Economics, University of Bonn.