
CORRELATED DEFAULT PROCESSES: A CRITERION-BASED COPULA APPROACH

Sanjiv R. Das^{a,*} and Gary Geng^b

In this paper, we develop a methodology to model, simulate and assess the joint default process of hundreds of issuers. Our study is based on a data set of default probabilities supplied by Moody's Risk Management Services. We undertake an empirical examination of the joint stochastic process of default risk over the period 1987 to 2000 using copula functions. To determine the appropriate choice of the joint default process, we propose a new metric. This metric accounts for different aspects of default correlation, namely (i) level, (ii) asymmetry and (iii) tail-dependence and extreme behavior. Our model, based on estimating a joint system of over 600 issuers, is designed to replicate the empirical joint distribution of defaults. A comparison of a jump model and a regime-switching model shows that the latter provides a better representation of the properties of correlated default. We also find that the skewed double-exponential distribution is the best choice for the marginal distribution of each issuer's hazard rate process, and combines well with the normal, Gumbel, Clayton and Student's t copulas in the joint dependence relationship amongst issuers. As a complement to the methodological innovation, we show that (a) appropriate choices of marginal distributions and copulas are essential in modeling correlated default, (b) accounting for regimes is an important aspect of joint specifications of default risk, and (c) misspecification of credit portfolio risk can occur easily if joint distributions are inappropriately chosen. The empirical evidence suggests that improvements are indeed possible over the standard Gaussian copula used in practice.



1 Introduction

Default risk at the level of an individual security has been extensively modeled using both structural and reduced form models.¹ This paper examines default

^aSanta Clara University, Santa Clara, CA 95053.

^bAmaranth Group Inc., Greenwich, CT 06831.

*Corresponding author. Santa Clara University, Santa Clara, CA 95053, USA. E-mail: srdas@scu.edu

risk at the portfolio level. Default dependences among issuers in a large portfolio play an important role in the quantification of a portfolio's credit risk exposure for many reasons. Growing linkages in the financial markets have led to a greater degree of joint default. While the actual portfolio loss due to the default of an individual obligor may be small unless the risk exposure is extremely large, the effects of simultaneous defaults of several issuers in a well diversified portfolio could be catastrophic. In order to efficiently manage and hedge the risk exposure to joint defaults, a model for default dependencies is called for. Further, innovations in the credit market have been growing at an unprecedented pace in recent years, and will likely persist in the near future. Many newly developed financial securities, such as collateralized debt obligations (CDOs), have payoffs depending on the joint default behavior of the underlying securities.² In order to accurately measure and price the risk exposure of these securities, an understanding and accurate measurement of the default dependencies among the underlying securities is essential.³ Accurate specification of default correlations is required in the pricing of basket default contracts. Correlation specifications are critical, as the baskets are not large enough to ensure diversification. Finally, the Basle Committee on Banking Supervision has identified poor portfolio risk management as an important source of risk for banking institutions. As a result, banks and other financial institutions have been required to measure their overall risk exposures on a routine basis. Default correlation is again an important ingredient to integrate multiple credit risk exposures.

In this paper, we use Moody's default database to develop a parsimonious numerical method of modeling and simulating correlated default processes for hundreds of issuers. Using copula functions, we capture the stylized facts of individual default probabilities as well as the dynamics of default correlations documented in Das *et al.* (2001a) (DFGK, 2001a). Our methodology is "criterion-based,"

i.e. it uses a metric that compares alternative specifications of the joint default distribution using three criteria: (a) the *level* of default risk, (b) the *asymmetry* in default correlations, and (c) the *tail-dependence* of joint defaults. We elaborate on these in what follows.

The base unit of joint credit risk is the probability of default (PD) of each issuer. Many popular approaches exist for computing PDs in the market place, developed by firms such as KMV Corporation, Moodys Risk Management Systems (MRMS), RiskMetrics, etc. In the spirit of reduced form models, the default probability of issuer i at time t is usually expressed as a stochastic hazard rate, denoted as $\lambda_i(t)$, $i = 1, \dots, N$. This paper empirically examines the *joint* stochastic process for hazard rates, $\lambda_i(t)$, for N issuers.

We fit a joint dynamic system to the intensities (or hazard rates) of roughly 600 issuers, classified into six rating categories. Our approach involves capturing the correlation of PD levels across the six rating classes, as well as the correlation of firms' PDs within each rating class. We find (a) the appropriate copula function (for the joint distribution), (b) the stochastic process for rating level default probabilities, and (c) the best set of marginal distributions for individual issuer default probabilities.

At the rating level, we choose from two classes of stochastic processes for the average hazard rate of the class: a jump-diffusion model, and a regime-switching one. At the issuer level, we choose from three distributional assumptions: a normal distribution, a Student's t distribution and a skewed double exponential distribution. We compare these distributions for each issuer using four goodness-of-fit criteria. At the copula level, we choose from four types: normal, Student's t , Gumbel and Clayton. These inject varying amounts of correlation emanating from the joint occurrence of extreme observations.⁴

The various combinations of choices at the rating, issuer and copula level result in 56 different systems of joint variation. To compare across these, we develop a *metric* to determine how well different multivariate distributions fit the observed covariation in PDs. The metric accounts for the level of correlations, asymmetry of correlations, and the tail-dependence in the data.

An important issue we look at here is “tail-dependence.” Tail-dependence is the feature of the joint distribution that determines how much of the correlation between default intensity processes comes from extreme observations than from central observations. For example, imagine two different bivariate distributions, with the same standard normal marginal distributions. One distribution has a normal copula and the other a Student’s t copula with a low degree of freedom. The former has lower tail-dependence than the latter, which generates more joint tail observations. Hence, from a risk point of view, given the marginal distributions, the second joint distribution would be riskier for a large class of risk-averse investors. Our study sheds light on the tail-dependence for default risk in a representative data set of US firms.

Our study of joint default risk for US corporates using copula techniques finds that the best choice of copula depends on the marginal distributions, and stochastic processes for rating level PDs. The Student’s t copula is best when the jump-diffusion model is chosen for the default intensity of each rating class. However, the normal or Clayton copula is more appropriate when the regime-switching model is chosen for rating level intensities. In general, our metric prefers a regime-switching model to a jump-diffusion one (for rating level intensities), irrespective of the choice of marginal distributions, and a perusal of the time series plot of the data (see Figure 1) provides intuitive support for this result. While the regime-switching model for rating-level intensities is best, no clear

winner emerges amongst the competing copulas. The interdependence between copula and marginal distributions justifies our wide-ranging search over 56 different specifications.

We also assess the impact of different copulas on risk-management measures. An examination of the tail loss distributions shows that substantial differences are possible amongst the 56 econometric specifications. Arbitrarily assigning high kurtosis distributions to a copula with tail-dependence may even result in the unnecessary overstatement of joint default risk. Since different copulas inject varied levels of tail-dependence, the metric developed in this paper allows fine-tuning of the specification, which enhances the accuracy of credit VaR calculations.

The rest of the paper proceeds as follows. In Section 2 we briefly describe the data on intensities. Section 3 provides the finance reader with a brief introduction to copulas, and Section 4 contains the estimation procedure and results. Section 5 presents the simulation model and the metric used to compare different copulas. Section 6 concludes.

2 Description of the data

Our data set comprises issuers tracked by Moodys from 1987 to 2000. For each issuer, we have PDs based on their econometric models for every month the issuer was tracked by Moodys. Moodys calibrates the PDs to match the level of realized defaults in the economy. Issuers are divided into seven rating classes. Rating classes 1 through 6 reflect progressively declining credit quality. Many more issuers fall into rating class 7, which comprises unrated issuers, and the PDs within this class range from high to low, resulting in an average PD that is close to the median PD of all the other rating

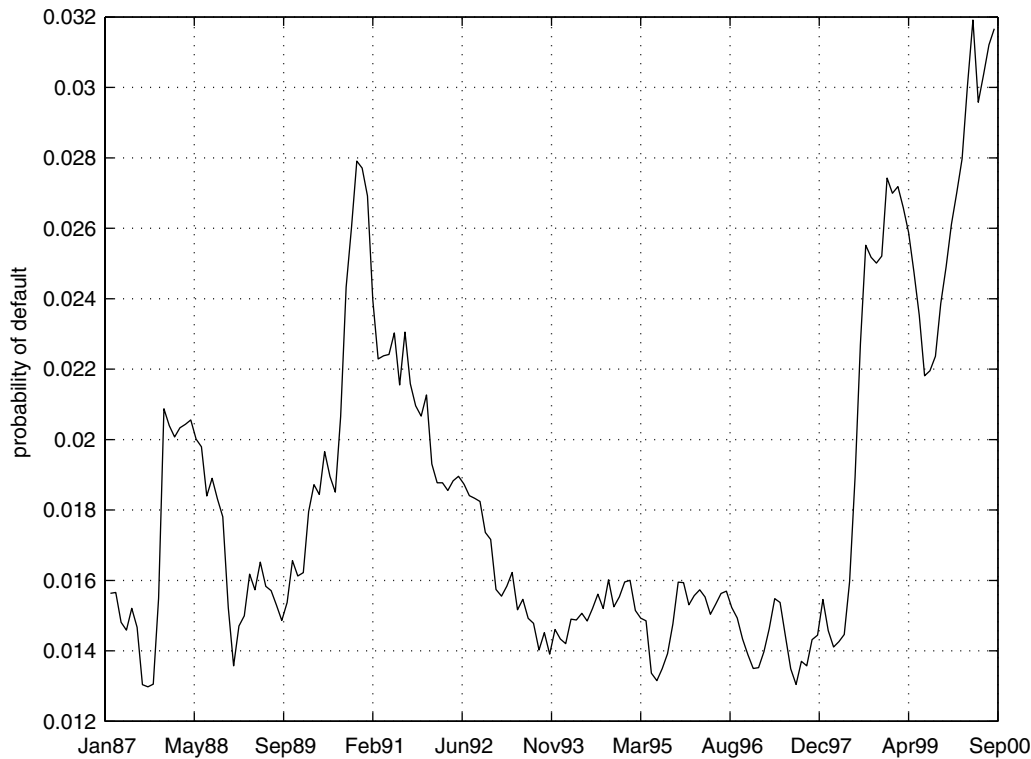


Figure 1 Time series of average PDs. This figure depicts the average level of default probabilities in the data set. The data shows the presence of two regimes, one in which PDs were high, as in the early and latter periods of the data. In the other regime, PDs were much lower, less than half of those seen in the high PD regime. Complementary analysis of correlated default on the same data set by subperiod in this same time frame is presented in DFGK (2001a) from where this figure is taken.

classes. We do not consider PDs from rating class 7. We considered firms that had a continuous sample over the period. We obtained data on a total of 620 issuers classified into six rating classes. These 1 year default probabilities were converted into intensities using an assumption of constant hazard rates, i.e. $\lambda_{it} = -\ln(1 - \text{PD}_{it})$. Table 1 provides summary data on Moody's rating categories.

The time series of average default probabilities is presented in Figure 1. Table 2 presents the descriptive statistics of our data from rating classes 1 through 6. The mean increases from rating class 1 to 6, as does the standard deviation. PD changes tend to be higher for lower grade debt. Table 3 presents the dependence between rating classes measured by Kendall's τ statistic. Higher grade debt

evidences more rank correlation in PDs than low grade debt. This highlights the fact that high grade debt is more systematically related, and low grade debt experiences greater idiosyncratic risk.

3 Copulas and features of the data

3.1 Definition

We are interested in modeling an n -variate distribution. A random draw from this distribution comprises a vector $X \in R^n = \{X_1, \dots, X_n\}$. Each of the n variates has its own marginal distribution, $F_i(X_i)$, $i = 1, \dots, n$. The joint distribution is denoted as $F(X)$. The copula associated with $F(X)$ is a multivariate distribution function defined

Table 1 Rating classes: the database comprises seven rating categories, in descending order of credit quality. Our data set does not contain data from firms in the financial sector.

-
- 1 Aaa/Aa: Bonds rated Aaa/Aa carry the smallest degree of investment risk and are generally known as high-grade bonds.
 - 2 A: Bonds rated A possess many favorable investment attributes and are to be considered as upper-medium-grade obligations
 - 3 Baa: Bonds rated Baa are considered as medium-grade obligations. Such bonds lack outstanding investment characteristics and in fact have speculative characteristics as well.
 - 4 Ba: Bonds rated Ba are judged to have speculative elements. Often the protection of interest and principal payments may be very moderate and thereby not well safeguarded during both good and bad times over the future.
 - 5 B: Bonds with B rating generally lack characteristics of the desirable investment. Assurance of interest and principal payments over any long period of time may be small.
 - 6 Caa/Ca/C: These are the bonds with high degree of speculation. Such issues may either be in default or elements of danger may be present with respect to principal or interest payments.
 - 7 Not rated.
-

Table 2 Descriptive statistics for PDs: This table presents the summary of the time series of average PDs for each rating class. Each time series represents a diversified portfolio of issuers within each rating class, equally weighted. The probabilities are in percentages.

Rating category	PD in levels		PD in changes	
	Mean	Std Dev.	Mean	Std Dev.
1	0.0728	0.1051	0.0012	0.0441
2	0.1701	0.4884	0.0036	0.1468
3	0.6113	1.6808	0.0044	0.7317
4	1.4855	2.6415	0.0066	1.2694
5	3.8494	4.6401	0.0138	2.3576
6	6.5389	4.7731	0.0165	3.2075

on the unit cube $[0, 1]^n$. When all the marginals, $F_i(X_i)$, $i = 1, \dots, n$, are continuous, the associated copula is unique and can be written as

$$C(u_1, \dots, u_n) = F(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)) \quad (1)$$

Every joint distribution may be written as a copula. This is Sklar's Theorem (Sklar, 1959, 1973). For an extensive discussion of copulas, see Nelsen (1999).⁵

Copulas allow the modeling of the marginal distributions separately from their dependence structure. This greatly simplifies the estimation problem of a joint stochastic process for a portfolio with many issuers. Instead of estimating all the distributional parameters simultaneously, we can estimate the marginal distributions separately from the joint distribution. Given the estimated marginal distribution for each issuer, we then use appropriate

Table 3 Kendall's τ for probabilities of default: In this table we report Kendall's τ , which is a measure of the dependence between any two time series. The upper right triangle below reports the measure for PD levels. The lower left triangle reports the correlation for PD changes.

Rating	Correlations					
	1	2	3	4	5	6
1	<u>1.0000</u>	0.6508	0.2724	0.2868	0.2987	0.2762
2	0.4254	<u>1.0000</u>	0.4671	0.4569	0.4052	0.1794
3	0.1849	0.2312	<u>1.0000</u>	0.6416	0.4483	0.1080
4	0.1447	0.1393	0.0543	<u>1.0000</u>	0.5382	0.1708
5	0.0378	0.0877	0.0352	0.1245	<u>1.0000</u>	0.2911
6	0.0362	-0.0429	0.0238	-0.0276	0.1032	<u>1.0000</u>

copulas to construct the joint distribution with a desired correlation structure. The best copula is determined by examining the statistical fit of different copulas to the data.⁶

Copula techniques lend themselves to two types of credit risk analysis. First, given a copula, we can choose different marginal distributions for each individual issuer. By changing the types of marginal distributions and their parameters, we can examine how individual default affects the joint default behavior of many issuers in a credit portfolio. Second, given marginal distributions, we can vary the correlation structures by choosing different copulas, or the same copula with different parameter values.

3.2 Copulas used in the paper

In this paper, in order to capture the observed properties of the joint default process, we consider the following four types of copulas:

Normal copula: The normal copula of the n -variate normal distribution with correlation matrix ρ , is defined as

$$C_{\rho}(u_1, \dots, u_n) = \Phi_{\rho}^n(F^{-1}(u_1), \dots, F^{-1}(u_n))$$

where $\Phi(u)$ is the normal cumulative distribution function (CDF), and $F^{-1}(u)$ is the inverse of the CDF.

Student's t copula: Let $T_{\rho, \nu}$ be the standardized multivariate Student's t distribution with ν degrees of freedom and correlation matrix ρ . The multi-variate Student's t copula is then defined as follows:

$$C(u_1, \dots, u_n; \rho, \nu) = T_{\rho, \nu}(t_{\nu}^{-1}(u_1), \dots, t_{\nu}^{-1}(u_n))$$

where t_{ν}^{-1} is the inverse of the cumulative distribution function of a univariate Student's t distribution with ν degree of freedom.

Gumbel copula: This copula was first introduced by Gumbel (1960) and can be expressed as follows:

$$C(u_1, \dots, u_n) = \exp \left[- \left(\sum_{i=1}^n (-\ln u_i)^{\alpha} \right)^{1/\alpha} \right] \quad (2)$$

where α is the parameter determining the tail-dependence of the distribution.

Clayton copula: This copula, introduced in Clayton (1978), is as follows:

$$C(u_1, \dots, u_n) = \left[\sum_{i=1}^n u_i^{-\alpha} - n + 1 \right]^{-1/\alpha} \quad (3)$$

Again, $\alpha > 1$ is a tail-dependence parameter.

The procedure for simulating an n -dimensional random vector from the above copulas can be found in Wang (2000), Bouye *et al.* (2000) and Embrechts *et al.* (2001).

3.3 Tail-dependence

An important feature of the use of copulas is that it permits varying degrees of *tail-dependence*. Tail-dependence refers to the extent to which the dependence (or correlation) between random variables arises from extreme observations.

Suppose (X_1, X_2) is a continuous random vector, with marginal distributions F_1 and F_2 . The coefficient of upper tail-dependence is:

$$\lambda_U = \lim_{z \rightarrow 1} Pr[X_2 > F_2^{-1}(z) \mid X_1 > F_1^{-1}(z)] \quad (4)$$

If $\lambda_U > 0$, then upper tail-dependence exists. Intuitively, upper tail-dependence is present when there is a positive probability of positive extreme observations occurring jointly.⁷

For example, the Gumbel copula has upper tail-dependence with $\lambda_U = 2 - 2^{1/\alpha}$. The Clayton copula has lower tail-dependence with $\lambda_L = 2^{-1/\alpha}$. The Student's t has equal upper and lower tail-dependence with $\lambda_U = 2\bar{t}_{v+1} \left(\frac{\sqrt{v+1}\sqrt{1-\rho}}{\sqrt{1+\rho}} \right)$.

Table 4 provides a snapshot of the tail-dependence between rating classes. The results in the table present the correlations between rating pairs over observations in the tails of the bivariate distribution.

Results are provided both, in levels and changes. There is strong positive correlation in the upper tail, evidencing upper tail-dependence. Correlations in the lower tail are often low and negative, and hence, there is not much evidence of lower tail-dependence.

3.4 Empirical features of dependence in the joint distribution

In order to compare the statistical fitting of joint distributions associated with different copulas, we develop a metric, which captures three features of the dependence relationship in the joint distribution:

- *Correlation levels*: We wish to ensure that our copula permits the empirically observed correlation levels in conjunction with the other moments. This depends on the copula and the attributes of the data.
- *Correlation asymmetry*: Hazard rate correlations are level dependent, and are higher when PD levels are high—correlations increase when PD levels jump up, and decrease when PDs jump down.
- *Tail-dependence*: It is important that we capture the correct degree of tail-dependence in the data, i.e. the extent to which extreme values drive correlations.

Following Ang and Chen (2002), and Longin and Solnik (2001), we present the correlations for different rating classes in a correlation diagram (see Figure 2). This plot is created as follows. We compute the total hazard rate (THR) across issuers (indexed by i) at each point in time (t), i.e. $THR_t = \sum_{i=1}^N \lambda_i(t)$, $\forall t$. We normalize the THR by subtracting the mean from each observation and dividing by the standard deviation. We then segment our data set based on exceedance

Table 4 Measures of tail-dependence between the different rating classes: this table presents correlations between hazard rates amongst the rating classes based on different tail percentiles. We examine correlation amongst observations in the bottom 10th, 20th, and 30th percentiles. We also present the correlations in the top 10th, 20th, and 30th percentiles, i.e. the cutoffs are the 70th, 80th, and 90th percentiles. The values in the table are the correlations when observations from the second rating in the pair of rating classes lies in the designated portion of the tail of the distribution. Results are presented for levels of hazard rates and for changes in hazard rates.

Rating Pair		Lower tail-dependence			Upper tail-dependence		
Rating	Rating	10%	20%	30%	70%	80%	90%
<i>Panel A: Tail-dependence in levels</i>							
1	2	0.4853	0.2879	0.2212	0.7518	0.7386	0.7206
1	3	0.0147	-0.3371	-0.1673	0.5837	0.5795	0.5441
1	4	-0.1471	0.0379	-0.0008	0.6376	0.4545	0.1471
1	5	-0.0882	-0.1439	-0.1314	0.6865	0.6061	0.2059
1	6	-0.2206	-0.1705	-0.0727	0.5298	0.4583	-0.0735
2	3	-0.1912	-0.2197	0.1004	0.5967	0.5985	0.5588
2	4	-0.2353	-0.1326	-0.0857	0.6000	0.4015	0.1324
2	5	-0.0147	-0.2045	-0.2065	0.6718	0.5189	0.0882
2	6	-0.1912	-0.3447	-0.1559	0.4171	0.3826	-0.1765
3	4	0.2500	0.2424	0.2539	0.5559	0.3902	0.1765
3	5	-0.0294	-0.1402	0.0694	0.5576	0.4205	0.2647
3	6	-0.1765	-0.1591	-0.0955	0.3535	0.4583	0.0147
4	5	0.0588	-0.1023	0.0155	0.7257	0.6439	0.4412
4	6	0.1471	-0.0833	-0.0318	0.4922	0.6326	0.4853
5	6	0.2206	0.0455	0.1020	0.4563	0.5152	0.2059
<i>Panel B: tail-dependence in Changes</i>							
1	2	0.5000	0.4394	0.4106	0.4253	0.4015	0.3971
1	3	0.0441	0.0189	0.0367	0.0776	0.0038	0.2206
1	4	0.1029	0.1439	0.0188	0.1429	0.0644	0.1618
1	5	0.1029	-0.0455	-0.0269	0.1069	0.1439	0.2794
1	6	0.0588	-0.0455	-0.0580	-0.2033	-0.1856	0.0735
2	3	0.1618	0.0606	0.0873	0.0204	-0.0606	0.1324
2	4	-0.2647	-0.0076	-0.0580	0.1984	0.0606	0.2647
2	5	-0.0147	0.0189	0.0645	0.0302	0.2045	0.3382
2	6	0.1618	0.0455	0.0482	-0.1559	-0.1667	-0.2500
3	4	0.2941	0.1364	-0.0433	0.0122	-0.0720	-0.0441
3	5	-0.0441	-0.1212	-0.0367	0.2082	-0.0152	0.0294
3	6	0.0882	0.0076	0.1216	-0.0694	-0.1136	-0.3382
4	5	-0.0588	-0.0227	0.1265	0.2555	0.3864	0.3824
4	6	-0.0735	0.1515	0.0971	0.0041	0.1098	-0.1324
5	6	-0.0294	0.1553	0.0171	0.0547	0.0341	-0.0147

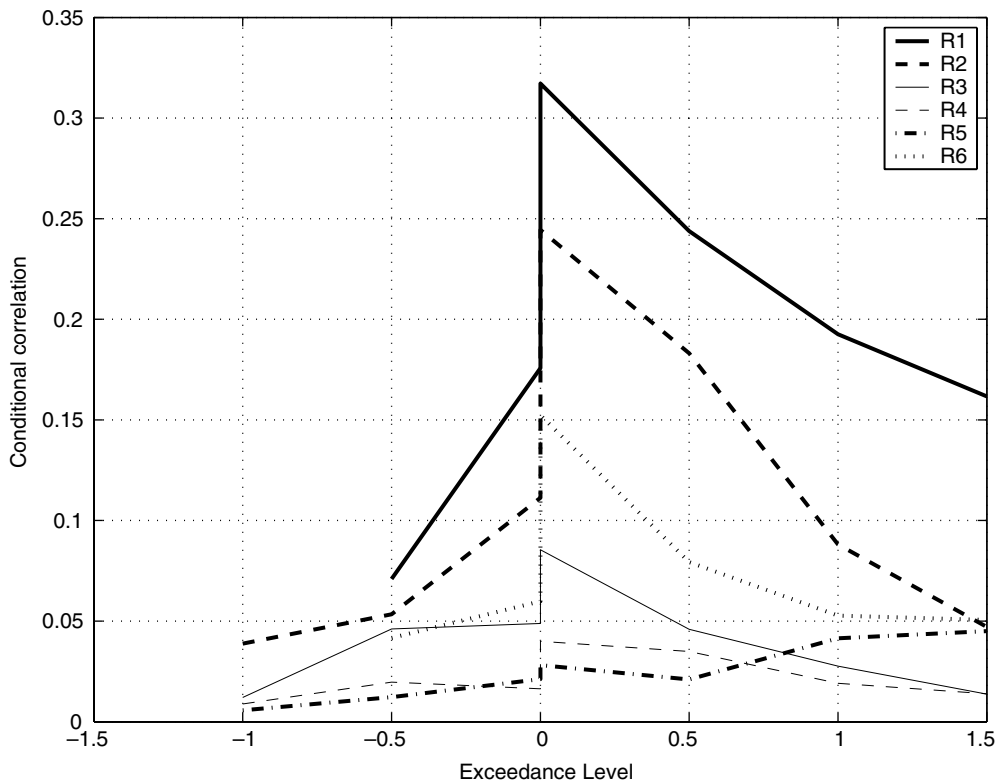


Figure 2 Asymmetric correlation of PDs: this figure plots the exceedance correlations for the different rating classes in our study. The figure breaks this out into the individual ratings. The plots are consistent with the presence of asymmetric correlation. The degree of asymmetry appears to be higher for the better rating categories. Complementary analysis of correlated default by subperiod in this same time frame is presented in DFGK (2001a) where different analyses are conducted on the same data set. This figure corresponds to their paper.

levels (ξ) determined by the normalized THR. Our exceedance levels are drawn from the set $\xi = \{-1.5, -1.0, -0.5, 0, 0, 0.5, 1.0, 1.5\}$. (Note that zero appears twice because we account for both, left and right exceedance.) Ang and Chen (2002) developed a version of this for bivariate processes. Our procedure here is a modification for the multivariate case. For example, the exceedance correlation at level $-\xi$ is determined by extracting the time series of PDs for which the normalized THR is less than $-\xi$, and computing the correlation matrix therefrom. We then find the average value of all entries in this correlation matrix, to obtain a summary number for the correlation in

the rating class at the given exceedance level. These are plotted in Figure 2. There is one line for each rating class. The line extends to both the left and right sides of the plot. To summarize, the exceedance graph shows (for each rating category) the correlation of default probabilities when the overall level of default risk exceeds different thresholds. As threshold levels become very small or very large, correlations decline, and the rate at which they decline indicates how much tail-dependence there is.

The three features of the dependence relationship are evident from the graph. First, the height of

the exceedance line indicates the level of correlation, and we can see that high grade debt has greater correlation than lower grade debt (see Xiao (2003) for a similar result with credit spreads). Second, there is clear evidence of correlation asymmetry, as correlation levels are much higher on the right side of the graph, i.e. when hazard rate changes are positive. Third, the amount of tail-dependence is inferred from the slope of the correlation line as the exceedance level increases. The flatter the line the greater the amount of tail-dependence. As absolute exceedance levels increase, we can see that correlation levels drop. However, they fall less slowly if there is greater tail-dependence. We can see that lower grade debt appears to have more tail-dependence than higher grade debt.

Overall, the exceedance plot shows that bonds within high quality ratings have greater default correlation than bonds within low quality ratings. However, tail-dependence is higher for lower rated bonds. Both results are intuitive—high grade debt tends to be issued by large firms that experience greater amounts of systematic risk, and low grade firms evidence more idiosyncratic risk; however, when economy-wide default risk escalates, low grade bonds are more likely to experience contagion, leading to greater tail-dependence.

4 Determining the joint default process

4.1 Overall method

Our data comprises firms categorized into six rating classes. We average across firms within rating class to obtain a time series of the average intensity (λ_k) for each rating class k . We assume that the stochastic processes for the six average λ_k s are drawn from a joint distribution characterized by a copula, which establishes the joint dependence between rating classes. The correlation matrix for

the mean intensities is calculated directly from the data. The copula is set to one of four possible choices: normal, Student's t , Gumbel or Clayton. The form of the stochastic process for each average hazard rate is taken to be either a jump-diffusion or a regime-switching one. This set up provides for the correlation *between* rating classes.

Next, the correlations *within* a rating class are obtained from the linkage of all firms in a rating class to the mean PD process of the rating category. We model individual firm PDs as following a non-negative stochastic process with reversion to the stochastic mean of each rating.

To conduct a Monte Carlo simulation of the system, we propagate the average rating PDs using either a jump-diffusion or regime-switching model, with correlations across ratings given by a copula. Then, we propagate the individual issuer intensities using stochastic processes that oscillate around the rating means. The entire system consists of a choice of (a) stochastic process for average intensities (jump-diffusion versus regime-switching), (b) copula (normal, Student's t , Gumbel, Clayton), and (c) marginal distribution for issuer-level stochastic process for intensities (one of seven different options). All told, therefore, we have 56 alternate correlated default system structures to choose from.

4.2 The estimation phase

Estimation is undertaken in a two-stage manner: (a) the estimation of the stochastic process for the mean intensity of each rating class, and (b) the estimation of the best candidate distribution for the stochastic process of each individual issuer's intensity. We chose two processes for step (a). Both choices were made to inject excess kurtosis into the intensity distribution. One, we used a normal-jump model. Two, we used a regime-switching

model. Each is described in turn in the next two subsections.

4.2.1 Estimation of the mean of each rating class using a jump model

For our six rating classes (indexed by k), and all issuers (indexed by j), we compute intensities from the probabilities of default (P_{kj} , $j = 1, \dots, N_k$, $k = 1, \dots, 6$) in our data set. The intensities are computed as: $\lambda_{kj} = -\ln(1 - P_{kj}) \geq 0$. We denote M as the total number of rating classes and N_k as the total number of issuers within the rating class for which data is available.

Let $\lambda_k(t)$ be the average intensity across issuers within rating class k . Therefore, $\lambda_k(t) = 1/N_k \sum_{j=1}^{N_k} \lambda_{kj}(t)$, $\forall t$. We assume that $\lambda_k(t)$ follows the stochastic process:

$$\begin{aligned} \Delta\lambda_k(t) &= \kappa_k[\theta_k - \lambda_k(t)]\Delta t \\ &\quad + x_k(t)\sqrt{\lambda_k(t)}\Delta t \\ x_k(t) &= \epsilon_k(t) + J_k(t)L_k(q_k, t) \\ \epsilon_k(t) &\sim N[0, \sigma_k^2] \\ J_k(t) &\sim U[a_k, b_k] \end{aligned} \quad (6)$$

$$L_k(q_k, t) = \begin{cases} 1 & \text{w/prob } q_k \\ 0 & \text{w/prob } 1 - q_k \end{cases}$$

We assume that the jump size J_k follows a uniform distribution over $[a_k, b_k]$.⁸ This regression accommodates a mean-reverting version of the stochastic process, and κ_k calibrates the persistence of the process. This process is estimated for each rating class k , and the parameters are used for subsequent simulation. Copulas are used to obtain the joint distribution of correlated default. The correlation matrix used in the copula comes from the residuals $x_k(t)$ computed in the regression above.

We use maximum-likelihood estimation to obtain the parameters of this process. Since there is a mixture of a normal and a jump component, we

can decompose the unconditional density function for the residual term $x_k(t)$ into the following:

$$f[x_k(t)] = q_k f[x_k(t) | L_k = 1] + (1 - q_k) f[x_k(t) | L_k = 0] \quad (7)$$

The latter density, $f[x_k(t) | L_k = 0]$ is conditionally normal. The marginal distribution of $x_k(t)$ conditional on $L_k = 1$ can be derived as follows:

$$\begin{aligned} f[x_k(t) | L_k = 1] &= \int_{a_k}^{b_k} \frac{1}{(b_k - a_k)\sqrt{2\pi\sigma_k^2}} \\ &\quad \times \exp\left\{-\frac{(x_k(t) - J_k)^2}{2\sigma_k^2}\right\} dJ_k \end{aligned} \quad (8)$$

Let $z = (J_k - x_k(t))/\sigma_k$. Then,

$$\begin{aligned} f[x_k(t) | L_k = 1] &= \int_{\frac{a_k - x_k(t)}{\sigma_k}}^{\frac{b_k - x_k(t)}{\sigma_k}} \frac{1}{b_k - a_k} \exp\left\{-\frac{z^2}{2}\right\} \frac{1}{\sqrt{2\pi}} dz \\ &= \frac{1}{b_k - a_k} \left\{ \Phi\left[\frac{b_k - x_k(t)}{\sigma_k}\right] - \Phi\left[\frac{a_k - x_k(t)}{\sigma_k}\right] \right\} \end{aligned}$$

In the case when $a_k = b_k$, the process has a constant jump size with probability 1 and the marginal distribution of $x_k(t)$ degenerates to a normal distribution with mean a_k and standard deviation σ_k . This can be seen by letting $b_k = a_k + c$. We, then, have

$$\begin{aligned} f[x_k(t) | L_k = 1] &= \lim_{c \rightarrow 0} \frac{1}{c} \left\{ \Phi\left(\frac{a_k + c - x_k(t)}{\sigma_k}\right) \right. \\ &\quad \left. - \Phi\left(\frac{a_k - x_k(t)}{\sigma_k}\right) \right\} \\ &= \lim_{c \rightarrow 0} \frac{1}{\sigma_k} \phi\left(\frac{a_k + c - x_k(t)}{\sigma_k}\right) \\ &= \frac{1}{\sigma_k} \phi\left(\frac{x_k(t) - a_k}{\sigma_k}\right) \end{aligned}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and distribution functions, respectively. Since

Table 5 Estimation of the average rating process: in this table we report the estimation results of the stochastic process for the mean of the intensities of each rating class. Numbers below the estimates are the t -statistics from the estimation, which is undertaken by maximum-likelihood.

Parameter	Rating class					
	1	2	3	4	5	6
θ_k	0.0524 7.73	0.1336 16.74	0.3499 2.75	1.1550 4.01	2.3812 1.99	5.4848 7.12
κ_k	0.4813 2.84	0.0018 0.01	0.6680 1.88	0.4811 1.87	0.4305 2.01	2.8350 4.19
σ_k	0.0487 14.17	0.1050 13.39	0.2518 12.48	0.2633 11.44	0.2872 8.51	0.5743 5.85
b_k	0.0839 12.39	0.1316 5.64	0.1241 1.94	0.2070 1.58	0.2684 1.52	1.1231 5.50
a_k	-0.0707 -6.33	-0.0644 -3.46	0.1241 1.96	0.2069 1.61	0.2684 2.37	0.0771 0.21
q_k	0.1016 3.87	0.0777 3.18	0.1514 2.94	0.0950 2.07	0.2461 2.08	0.4374 2.07
Log-likelihood	-622.16	-462.95	-205.48	-133.56	-32.82	-35.03

jumps are permitted to be of any sign, we may expect that $b_k \geq 0$ and $a_k \leq 0$. Also, if $|b_k| > |a_k|$, it implies a greater probability of positive jumps. When a_k is close to b_k , or when they have the same sign, the process has a constant or one-direction jump, respectively.

The estimation results are presented in Table 5. The mean of the intensity process (θ_k) increases with rating class k as expected. The variance (σ_k) also increases with declining credit quality. Generally speaking, the probability of a jump (q_k) in the hazard rate is higher for lower quality issuers. We can see that rating classes 3–5 have a constant jump and rating 6 has only positive jumps. While the mean jump for all rating classes appears to be close to zero, that for the poorest rating class is much higher than zero. Hazard rates jump drastically when a firm approaches default.

After obtaining the parameters, we compute residuals for each rating class. The residuals include randomness from both the normal and jump terms. The covariance matrix of these residuals is stored for later use in Monte Carlo simulations.

4.2.2 Estimating the mean processes in a regime-switching environment

From Figure 1, we note that there are periods in which PDs are low, interjected by smaller, sporadic regimes of spikes in the hazard rates. A natural approach to capture this behavior is to use a regime-switching model.⁹

In order to determine regimes, we first computed the average intensity (hazard rate, approximately) $\bar{\lambda} = \sum_{i=1}^N \lambda_i$ across all issuers in our database. Within each regime the intensity is assumed to follow a square-root volatility model represented in

discrete-time as follows:

$$\Delta \bar{\lambda}_r(t) = \kappa_r [\theta_r - \bar{\lambda}_r(t)] \Delta t + \sigma_r \sqrt{\bar{\lambda}_r(t)} \Delta t \epsilon(t),$$

$$r = \{HI, LO\} \quad (9)$$

The two regimes are indexed by r , which is either HI or LO. κ_r is the rate of mean-reversion. The mean value within the regime is θ_r and σ_r is the volatility parameter.

The probability of switching between regimes comes from a logit model based on a transition matrix:

$$\begin{bmatrix} p_{LO} & 1 - p_{LO} \\ 1 - p_{HI} & p_{HI} \end{bmatrix}$$

where $p_r = \exp(\alpha_r) / (1 + \exp(\alpha_r))$, $r \in \{LO, HI\}$. Estimation is undertaken using maximum-likelihood. We fixed the values of θ_r based on the

historical PDs, with a high–low regime cutoff of 2%. θ_{LO} was found to be 1.68% and θ_{HI} to be 2.40%.¹⁰ We then estimated the rest of the parameters in the above model. The estimation results are presented in Table 6. In the high PD regime, the higher level of hazard rates is matched by a higher level of volatility.

Our next step is to use the determined regimes to estimate the parameters of the mean process for each rating class, i.e. λ_k , $k = 1, \dots, 6$. We fit regime-shifting models to each of the rating classes, using a stochastic process similar to that in Eq. (9):

$$\Delta \bar{\lambda}_{k,r}(t) = \kappa_{k,r} [\theta_{k,r} - \bar{\lambda}_{k,r}(t)] \Delta t$$

$$+ \sigma_{k,r} \sqrt{\bar{\lambda}_{k,r}(t)} \Delta t \epsilon(k, t),$$

$$r = \{HI, LO\}, \quad k = 1, \dots, 6. \quad (10)$$

Table 6 Estimation of the regime-switching model across all issuers: this table provides estimation results for regimes estimated on the average intensity process ($\bar{\lambda} = \sum_{i=1}^N \lambda_i$) across all issuers in the data set. We designated periods where $\bar{\lambda} \leq 2\%$ as the “low-PD” regime, and periods in which $\bar{\lambda} > 2\%$ as the “high-PD” regime. The two regimes are indexed by r , which is either HI or LO. κ_r is the rate of mean-reversion. The mean value within the regime is θ_r and σ_r is the volatility parameter. The regimes were bifurcated exogenously, as described, and we estimated the values of θ_r by simply averaging the hazard rate within each regime. The probability of switching between regimes comes from a logit model based on a transition matrix: $\begin{bmatrix} p_{LO} & 1-p_{LO} \\ 1-p_{HI} & p_{HI} \end{bmatrix}$, where $p_r = \frac{\exp(\alpha_r)}{1 + \exp(\alpha_r)}$, $r \in \{LO, HI\}$. Estimation is undertaken using maximum-likelihood.

Parameter	Low-PD regime	High-PD regime
θ	1.68%	2.40%
κ	1.5043	1.7746
t -stat	2.4053	2.0558
σ	0.1681	0.2397
t -stat	14.2093	7.9849
α	4.1286	3.3644
t -stat	4.6649	2.5428
Loglik	-195.5948	

Table 7 Estimation of the regime-switching model for the mean of each rating class: this table presents parameters for the regime-switching model applied to the mean process for each rating class, i.e. $\bar{\lambda}_k$, $k = 1, \dots, 6$.

Parameter	Rating class					
	1	2	3	4	5	6
θ_{LO}	0.0559	0.1625	0.5678	1.3588	3.6367	6.3138
t -stat	0.40	7.06	11.90	14.22	45.24	74.32
θ_{HI}	0.0991	0.3757	0.9818	2.2358	5.1216	7.8295
t -stat	0.36	6.55	4.91	19.29	18.74	8.07
κ_{LO}	1.3535	0.5094	1.5855	1.1317	2.2365	6.2456
t -stat	0.23	1.31	2.86	2.25	3.49	6.45
κ_{HI}	0.6152	0.1775	1.1717	2.8567	2.0833	1.8862
t -stat	0.35	1.74	0.95	2.52	1.67	1.81
σ_{LO}	0.0568	0.1224	0.3214	0.2986	0.3338	0.6798
t -stat	17.63	17.27	16.19	16.57	16.92	16.02
σ_{HI}	0.1103	0.2214	0.3275	0.3305	0.3681	0.8336
t -stat	8.41	7.83	7.84	7.93	8.15	8.24

The results of the estimation are presented in Table 7. The parameters are as expected. The mean of the processes is higher in the high regime, as is the volatility. We now proceed to look at the choice of individual issuer marginal distributions.

4.2.3 Estimation of the stochastic process for each individual issuer

Assume that the intensity for each individual issuer follows a square-root process:

$$\begin{aligned} \Delta\lambda_{kj}(t) = & \kappa_{kj}\{\theta_{kj} + \gamma_{kj}\lambda_k - \lambda_{kj}(t)\} \\ & + \sigma_{kj}\sqrt{\lambda_{kj}(t)}\varepsilon_{kj}(t) \end{aligned} \quad (11)$$

We normalize the residuals by dividing both sides by $\sqrt{\lambda_{kj}(t)dt}$. We then use OLS to estimate parameters, which are stored for use in simulations. The resulting residuals are used to form the covariance matrix for each class.

The long-term mean is defined as the sum of a constant number θ_{kj} , and the group mean $\lambda_k(t)$ multiplied by a constant, γ_{kj} . This indexes the mean hazard rate for each issuer to the current value of the mean for the rating class. Consequently, there are four parameters for each regression, κ_{kj} , θ_{kj} , γ_{kj} and σ_{kj} . Note that κ_{kj} should be positive, θ_{kj} and γ_{kj} can be positive or negative as long as they are not both negative at the same time. None of the issuers have negative θ_{kj} and γ_{kj} at the same time, although 20 issuers have negative κ_{kj} . We, therefore, deleted the 20 issuers with negative κ from our dataset for further analysis.¹¹

4.2.4 Estimation of the marginal distributions

We fitted the residuals from the previous section to the normal, Student's t , and skewed double-exponential distributions. We used prepackaged functions in Matlab for maximum-likelihood estimation for the normal and Student's t distributions. The likelihood function for the skewed

double exponential distribution is as follows. Assume that the residual ε_{kj} has a normal distribution with mean γV and variance V . Further, the variance V is assumed to have an exponential distribution with the following density function:

$$\text{pdf}(V) = \frac{1}{V_0} \exp\left(-\frac{V}{V_0}\right) \quad (12)$$

Then, it can be shown that ε_{kj} has a skewed double-exponential distribution with the following density function:

$$\text{pdf}(\varepsilon) = \frac{\lambda}{V_0} \exp\left(-\frac{|\varepsilon_{kj}|}{\lambda} - \gamma \varepsilon_{kj}\right), \quad (13)$$

where

$$\lambda = \sqrt{\frac{V_0}{2 + \gamma^2 V_0}}$$

See the Appendix for a derivation of these results. The log-likelihood function L is given as follows:

$$L = -\frac{n}{2} \{ \log(V_0) + \log(2 + \gamma^2 V_0) \} \\ - \sum_{i=1}^n |\varepsilon_i| \sqrt{\frac{2 + \gamma^2 V_0}{V_0}} + \gamma \sum_{i=1}^n \varepsilon_i.$$

4.2.5 Goodness of fit of marginal distributions

Since we choose one of three distributions for the residuals of each issuer, we used four criteria to decide the best one. These are as follows:

- The Kolmogorov distance: this is defined as the supremum over the absolute differences between two cumulative density functions, the empirical one, $F_{\text{emp}}(x)$ and the estimated, fitted one, $F_{\text{est}}(x)$.
- The Anderson and Darling Statistic: this is given by

$$AD = \max_{x \in R} \frac{|F_{\text{emp}}(x) - F_{\text{est}}(x)|}{\sqrt{F_{\text{est}}(x)[1 - F_{\text{est}}(x)]}}$$

The AD statistic puts more weight on the tails compared to the Kolmogorov statistic.

- The L^1 distance: equal to the average of the absolute differences between the empirical and statistical distributions.
- The L^2 distance: which is the root-mean-squared difference between the two distributions.

We used these criteria to choose between the normal, Student's t and skewed double-exponential distributions for each issuer. In each of our 56 simulated systems, the issuer level distributions are chosen from the following seven cases:

1. All marginal distributions were chosen to be normal.
2. All marginal distributions are Student's t .
3. All marginals are from the skewed double-exponential family.
4. For each issuer, the best distribution is chosen based on the Kolmogorov criterion.
5. For each issuer, the best distribution is chosen based on the Anderson and Darling statistic.
6. Each marginal is chosen based on the best L^1 distance.
7. Each marginal is chosen based on the best L^2 distance.

For each criterion, we counted the number of times each distribution provided the best fit. The results are reported in Table 8, which provides interesting features. Based on the Kolmogorov, Anderson–Darling, and L^2 statistics, the skewed double-exponential distribution is the most likely to fit the marginals best. However, the L^1 statistic finds that the normal distribution fits most issuers better. The better fit of the skewed double-exponential as marginal distribution comes from its ability to match the excess kurtosis of the distribution.

4.3 Calibration

Our calibration procedure is as follows. First, using MLE, we determine the parameters of the various

Table 8 Fitting marginal distributions to individual intensities: in this table we report the estimation results of the fit of individual intensity residuals to various distributions. The three distributions chosen were: double exponential, normal, and the Student's t . We used four distance metrics to compare the empirical residuals to standardized distributions: the Kolmogorov distance, the Anderson–Darling statistic, and distances in the L^1 and L^2 norms. The table below presents the number of individual issuers that best fit each of the distributions under the different metrics. A total of 619 issuers was classified in this way.

Statistic	Double exponential	Normal	Student's t
Kolmogorov	522	55	42
Anderson–Darling	409	7	203
L^2	529	53	37
L^1	118	500	1

stochastic processes (rating level averages and individual issuer PDs). Second, we choose one of the 56 possible correlation systems as described above, and simulate the entire system 100 times, to obtain an average simulated exceedance graph. Third, we compare the simulated exceedance graph with the empirical one (pointwise) and determine the mean squared difference between the two graphs. This distance metric is used to determine a ranking of the 56 possible systems we work with.

5 Simulating correlated defaults and model comparisons

5.1 Overview

In this section, we discuss the simulation procedure. This step uses the estimated parameters from the previous section to generate correlated samples of hazard rates. The goal is for the simulation approach to deliver a model with the three properties described in the previous section. Therefore, the asymmetric correlation plot from simulated data should be similar to that in Figure 2.

We implemented the simulation model with an additional constraint, whereby we ensure that the

intensities ($\lambda_k(t)$) are monotonically increasing in rating level k . This prevents the average PD for a rating class from being lower than the average PD for the next better rating category, i.e. $\lambda_k(t)$ should be such that for all time periods t , it must be such that if $i < j$, then $\lambda_i(t) < \lambda_j(t)$. This check was instituted during the simulation as follows. During the Monte Carlo step, if $\lambda_i(t) > \lambda_j(t)$ when $i < j$, then we set $\lambda_j(t) = \lambda_i(t)$. As an example, see Figure 3 where we plot the times series of $\lambda_k, k = 1, \dots, 6$ for a random simulation of the sample path of hazard rates.

5.2 Illustrative Monte Carlo experiment

As a simple illustration that our approach arrives at a correlation plot fairly similar to that seen in the data, we ran a naive Monte Carlo experiment. In this exercise, we assumed that all error terms were normal, except under some conditions, when we assumed the Student's t distribution. We ran the Monte Carlo model 25 times and computed the average exceedance correlations for all rating classes. To obtain asymmetry in the correlations, we need to have higher correlations when hazard rates are high. To achieve this, the simulation uses the Student's t distribution with six



Figure 3 Simulated series of average PDs by rating class: this figure depicts the average level of default probabilities in the data set by rating class. The hazard rates are (from top to bottom of the graph) from rating class six to rating class one. The simulation ensures that $\lambda_i(t) < \lambda_j(t)$ if $i < j$.

degrees of freedom when the average level of hazard rates in the previous period in the simulation is above the empirical average of hazard rates. These features provided the results in Figure 4. The similarity between the empirical exceedance graph (Figure 2) and the simulated one (Figure 4) shows that the two-stage Monte Carlo model is able to achieve the three correlation properties of interest.

5.3 Determining goodness of fit

The dependence structure amongst the intensities is depicted in the asymmetric correlation plots. We develop a measure to assess different specifications of the joint distribution. This metric is the average

squared point-wise difference between the empirical exceedance correlation plot and the simulated one. The points in each plot that are used are for the combination of rating class (from 1 to 6) and exceedance levels (from -1.5 to $+1.5$). Hence, there are a total of 48 points in each plot which are used for computing the metric. Define the points in the empirical plot as $h_{k,x}$, where k indexes the rating class and x indexes the exceedance levels. The corresponding points in the simulated plot are denoted by $h'_{k,x}$. The metric q (a RMSE statistic) is as follows:

$$q = \sqrt{\frac{1}{48} \sum_{k=1}^6 \sum_{x=-1.5}^{1.5} (h_{k,x} - h'_{k,x})^2} \quad (14)$$

A smaller value of q implies a better fit of the joint dependence relationship. In the following section,

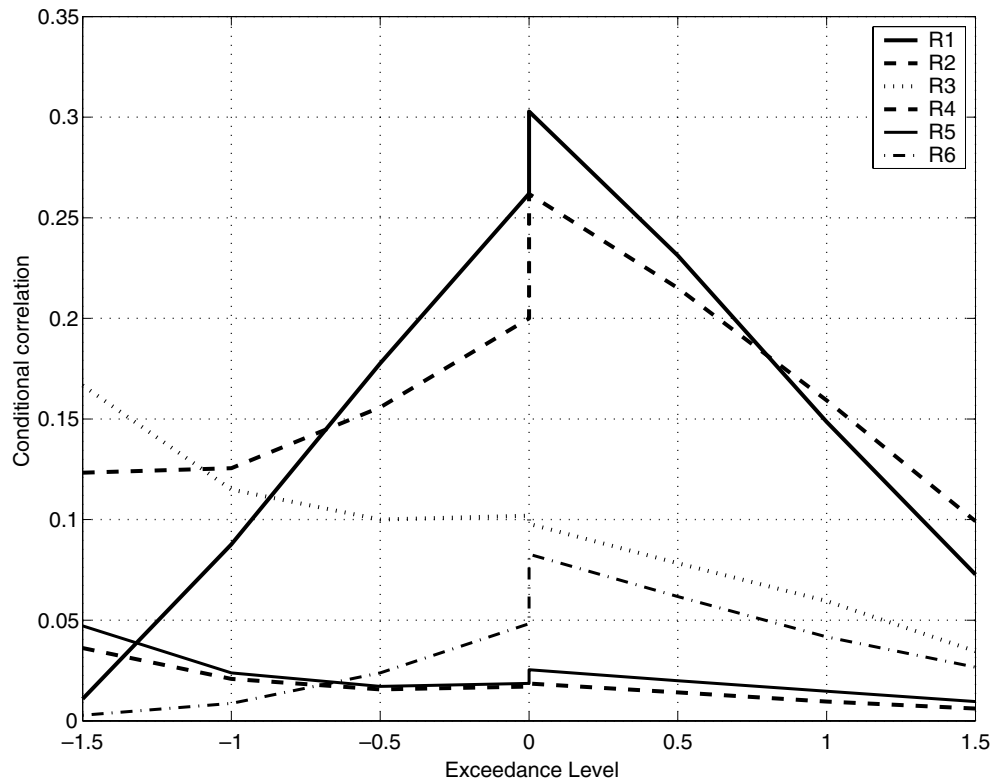


Figure 4 Asymmetric correlation of simulated PDs: this figure plots the exceedance correlations for the different rating classes in our study, based on simulated data. The figure breaks this out into the individual ratings. The plots are consistent with the presence of asymmetric correlation. The degree of asymmetry is higher for the better rating categories.

we use the simulation approach and this metric to compare various models of correlated default.

5.4 Calibration results and metric

It is important to note that the joint dependence across all issuer intensities depends on three aspects: (a) the marginal distributions used for individual issuers, (b) the stochastic process used for each rating class, and (c) the copula used to implement the correlations between rating classes. Optimizing the choice of individual issuer marginal distributions will not necessarily achieve the best dependence relationship, since the copula chosen must also be compatible. In this paper, we focus on four copulas (normal, Student's t , Gumbel, and Clayton),

and allow the rating class intensity stochastic process to be either a jump-diffusion model or a regime-shifting model.

Tables 9 and 10 provide a comparison of the jump and the regime-switching models across copulas. Twenty-eight systems use a jump-diffusion model for the average PDs in each rating, and 28 use a regime-switching model. We see that the values of the q metric are consistently smaller for the regime-switching model using all test statistics. This implies that the latter model provides a better representation of the stochastic properties amongst the issuer hazard rates. In particular, it fits the asymmetric correlations better.¹² We conclude that, for rating level PDs, a regime-switching model

Table 9 Metric for best correlated default model (normal and Gumbel copulas): this table presents the summary statistic for the asymmetric correlation metric to determine the best simulation model. For each model we generate the asymmetric correlation plot and then compute the distance metric. The asymmetric correlations (the metric q) are computed for the following seven cases: normal distribution, Student- t distribution, skewed double-exponential distribution, the combination of the best distributions based on Kolmogorov criterion, the combination of the best distributions based on the Anderson and Darling statistic, the combination based on the L^1 and L^2 norms. Our metric comprises the average squared point-wise difference between the empirical exceedance correlation plot and the simulated one. This a natural distance metric. The points in each plot that are used are for the combination of rating class (from 1 to 6) and exceedance levels (from -1.5 to $+1.5$). Hence, there are a total of 48 points in each plot which are used for computing the metric. Define the points in the empirical plot as $h_{k,x}$, where k indexes the rating class and x indexes the exceedance levels. The corresponding points in the simulated plot are denoted $h'_{k,x}$.

The metric q (a RMSE statistic) is as follows: $q = \sqrt{\frac{1}{48} \sum_{k=1}^6 \sum_{x=-1.5}^{1.5} [h_{k,x} - h'_{k,x}]^2}$. We report the results for both models, the jump-diffusion setup and the regime-switching one, and two copulas, the Gaussian and Gumbel copulas.

Model	Gaussian copula		Gumbel copula	
	Jump-diffusion	Regime-shifting	Jump-diffusion	Regime-shifting
Normal	0.0508	0.0276	0.0486	0.0462
Student's t	0.0375	0.0327	0.0416	0.0515
Double-exponential	0.0752	0.0355	0.0700	0.0381
Kolmogorov	0.0802	0.0387	0.0743	0.0364
Anderson–Darling	0.0447	0.0315	0.0452	0.0450
L^2	0.0764	0.0367	0.0710	0.0376
L^1	0.0667	0.0318	0.0625	0.0401

performs better than a jump-diffusion model, irrespective of the choice of issuer process and copula.¹³

Tables 9 and 10 allows us to compare the fit provided by the four copulas. We find that when the jump-diffusion model is used, the Student's t copula performs best, no matter which marginal distribution is used. However, when the regime-switching model is used, sometimes the normal copula works best and at other times, the Clayton copula is better. The normal copula works best when the marginals are all normal or Student's t , and when the marginals are chosen using the Anderson–Darling criterion. The Clayton copula is best when

the marginals are all skewed double-exponential or when the marginal criteria are Kolmogorov or L^2 . The Student's t copula is best when the L^1 criterion is chosen.

These results bear some explanation. A comparison of the correlations from the skewed double-exponential distribution with those from the raw data reveals that the model overestimates the up-tail dependences a lot. The Clayton copula decreases the up-tail dependences and increases the low-tail dependences. As a result, it corrects some of the over-estimation from the double-exponential marginals, resulting in better fitting. Hence, the Clayton copula seems to combine well

Table 10 Metric for best correlated default model (Clayton and Student's t copulas): this table presents the summary statistic for the asymmetric correlation metric to determine the best simulation model. For each model we generate the asymmetric correlation plot and then compute the distance metric. The asymmetric correlations (the metric q) are computed for the following seven cases: normal distribution, Student- t distribution, skewed double-exponential distribution, the combination of the best distributions based on Kolmogorov criterion, the combination of the best distributions based on the Anderson and Darling statistic, the combination based on the L^1 and L^2 norms. Our metric comprises the average squared point-wise difference between the empirical exceedance correlation plot and the simulated one. This a natural distance metric. The points in each plot that are used are for the combination of rating class (from 1 to 6) and exceedance levels (from -1.5 to $+1.5$). Hence there are a total of 48 points in each plot which are used for computing the metric. Define the points in the empirical plot as $h_{k,x}$, where k indexes the rating class and x indexes the exceedance levels. The corresponding points in the simulated plot are denoted $h'_{k,x}$. The metric q (a RMSE statistic) is as follows: $q = \sqrt{\frac{1}{48} \sum_{k=1}^6 \sum_{x=-1.5}^{1.5} [h_{k,x} - h'_{k,x}]^2}$. We report the results for both models, the jump-diffusion set up and the regime-switching one, and two copulas, the Clayton and Student's t copulas.

Model	Clayton copula		Student's t copula	
	Jump-diffusion	Regime-shifting	Jump-diffusion	Regime-shifting
Normal	0.0484	0.0395	0.0446	0.0506
Student's t	0.0415	0.0473	0.0370	0.0564
Double-exponential	0.0694	0.0304	0.0672	0.0433
Kolmogorov	0.0738	0.0293	0.0714	0.0404
Anderson–Darling	0.0445	0.0397	0.0418	0.0514
L^2	0.0703	0.0296	0.0681	0.0423
L^1	0.0622	0.0324	0.0596	0.0452

with skewed double-exponential marginals. On the other hand, the q -metric favors the normal copula if the criterion chooses normal marginals (as with the normal, Student's t and L^1 criteria), since less balancing of tail-dependences is required.

5.5 Empirical implications of different copulas

The degree of tail-dependence varies with the choice of copula. It is ultimately an empirical question as to whether the parameterized joint distribution does result in differing tail risk in credit

portfolios. To explore this question, we simulated defaults using a portfolio comprising all the issuers in this study, under the regime-switching model. Remember that we fitted four copulas, and seven different choices of marginal distributions, resulting in 28 different models, each with its attendant parameter set. To compare copulas, we fix the marginal distribution, and then vary the copula.

As an illustration, we present Figure 5. This plots the tail loss distributions from the four copulas when all marginals are assumed to be normal. The line most to the right (the Gumbel copula) has the most tail-dependence. There is roughly a 90%

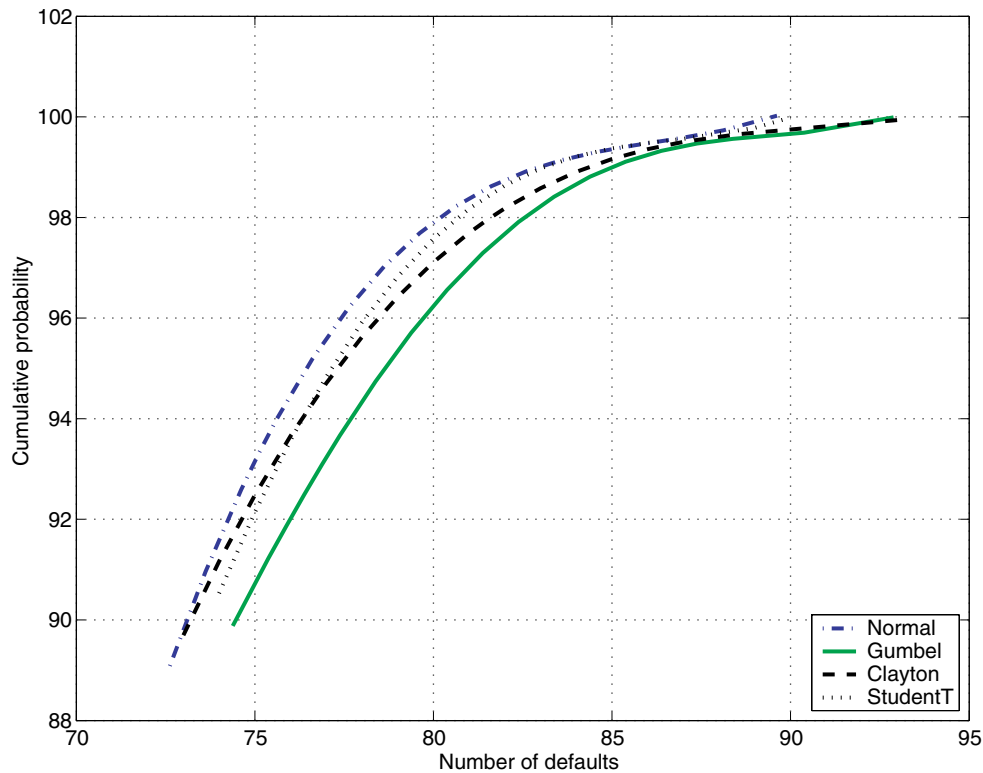


Figure 5 Comparing copula tail loss distributions: this figure presents plots of the tail loss distributions for four copulas, when the marginal distribution is normal. The x -axis shows the number of losses out of more than 600 issuers, and the y -axis depicts the percentiles of the loss distribution. The simulation runs over a horizon of 5 years and accounts for regime shifts as well. The copulas used are: normal, Gumbel, Clayton, Student's t .

chance that the number of defaults will be less than 75, i.e. a 10% probability that the number of defaults will exceed 75. For the same level of 75 defaults, the leftmost line (from the normal copula), there is only a 7% probability of the number of losses being greater than 75. By examining all four panels of the figure, we see that the ranking of copulas by tail-dependence is unaffected by the choice of marginal distribution, i.e. the normal copula is the leftmost plot, followed by the Student's t , Clayton, and Gumbel copulas.

In Figure 6 we plot the tail loss distributions for two models, the best fitting one and the worst. The best fit model combines the Clayton copula, and

marginal distributions based on the Kolmogorov criterion. The worst fit copula combines the Student's t copula with Student's t marginals. A comparison of the two models shows that the worst fitted copula in fact grossly overstates the extent of tail loss. Therefore, while there is tail-dependence in the data, careful choice of copula and marginals is needed to avoid either under- or over-estimation of tail-dependence.

6 Discussion

This paper develops a criterion-based methodology to assess alternative specifications of the joint distribution of default risk across hundreds of issuers

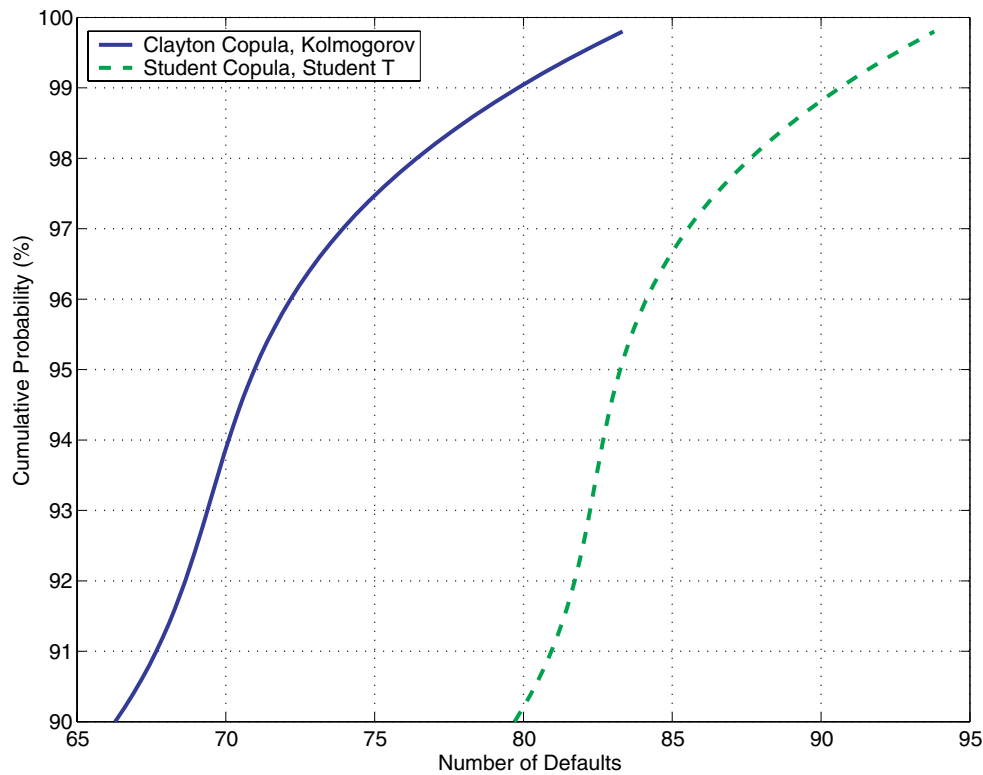


Figure 6 Comparing copula tail loss distributions: in this figure we plot the tail loss distributions for two models, the best fitting one and the worst. The best fit model combines the Clayton copula, and marginal distributions based on the Kolmogorov criterion. The worst fit copula combines the Student's t copula with Student's t marginals. The simulation runs over a horizon of 5 years and accounts for regime shifts as well.

in the US corporate market. The study is based on a data set of default probabilities supplied by Moody's Risk Management Services. We undertake an empirical examination of the joint stochastic process of default risk over the period 1987–2000. Using copula functions, we separate the estimation of the marginal distributions from the estimation of the joint distribution. Using a two-step Monte Carlo model, we determine the appropriate choice of multivariate distribution based on a new metric for the assessment of joint distributions.

We explored 56 different specifications for the joint distribution of default intensities. Our methodology uses two alternative specifications (jumps and regimes) for the means of default rates in rating

classes. We consider three marginal distributions for individual issuer hazard rates, combined using four different copulas. An important extension to this model structure is the inclusion of rating changes. In our analysis, we centered firms as lying within the same rating for the period of the simulation, based on their most prevalent rating. The two-step simulation model would need to be enhanced to a three-step one, with an additional step for changes in ratings.¹⁴

Other than the myriad specifications, there are many useful features of the analysis for modelers of portfolio credit risk. First, we developed a simple metric to measure best fit of the joint default process. This metric accounts for different aspects

of default correlation, namely level, asymmetry and tail-dependence or extreme behavior. Second, the simulation model, based on estimating the joint system of over 600 issuers, is able to replicate the empirical joint distribution of default. Third, a comparison of the jump model and the regime-switching model shows that the latter provides a better representation of the properties of correlated default. Fourth, the skewed double exponential distribution is a suitable choice for the marginal distribution of each issuer hazard rate process, and combines well with the Clayton copula in the joint dependence relationship amongst issuers. Our simulation approach is fast and robust, allowing for rapid generation of scenarios to assess risk in credit portfolios. Finally, the results show that it is important to correctly capture the interdependence of marginal distributions and copula to achieve the best joint distribution depicting correlated default. Thus, this paper delivers the empirical counterpart to the body of theoretical papers advocating the usage of copulas in modeling correlated default.

Acknowledgments

We are extremely thankful for many constructive suggestions and illuminating discussions with Darrell Duffie, Gifford Fong, Nikunj Kapadia, John Knight, and Ken Singleton, and participants at the Q-conference (Arizona 2001), the Risk Conference (Boston 2002), and at Barclays Global Investors Seminar (San Francisco 2002). We are also very grateful to two excellent (anonymous) prior referees for superb comments and guidance, which resulted in a much-transformed paper. The first author gratefully acknowledges support from the Dean Witter Foundation, and a Breetwor Fellowship. We are also grateful to Gifford Fong Associates, and Moody's Investors Services for data and research support for this paper. The second author is supported by the Natural Sciences and Engineering Research Council of Canada.

Appendix: The skewed double-exponential distribution

Assume that a random variable X has a normal distribution with mean $\mu + \gamma V$, and variance V , where V has an exponential distribution with the following density function:

$$\text{pdf}(V) = \frac{1}{V_0} \exp\left(-\frac{V}{V_0}\right) \quad (15)$$

Then, X has a skewed double-exponential distribution.

A.1 Density function

The density function is derived as follows:

$$\begin{aligned} \text{pdf}(x) &= \int_0^\infty \frac{1}{\sqrt{2\pi V}} \exp\left\{-\frac{(x - \mu - \gamma V)^2}{2V}\right\} \\ &\quad \times \frac{1}{V_0} \exp\left(-\frac{V}{V_0}\right) dV \\ &= \frac{1}{\sqrt{2\pi} V_0} \exp\{\gamma(x - \mu)\} \int_0^\infty \frac{1}{\sqrt{V}} \\ &\quad \times \exp\left\{-\frac{(x - \mu)^2}{2V} - \frac{\gamma^2 V_0 + 2}{2V_0} V\right\} dV \\ &= \frac{1}{\sqrt{2\pi} V_0} \exp\{\gamma(x - \mu)\} \sqrt{\frac{2\pi V_0}{2 + \gamma^2 V_0}} \\ &\quad \times \exp\left\{-|x - \mu| \sqrt{\frac{2 + \gamma^2 V_0}{V_0}}\right\} \\ &= \frac{1}{V_0} \sqrt{\frac{V_0}{2 + \gamma^2 V_0}} \\ &\quad \times \exp\left\{-|x - \mu| \sqrt{\frac{2 + \gamma^2 V_0}{V_0}}\right. \\ &\quad \left. + \gamma(x - \mu)\right\} \\ &= \frac{\lambda}{V_0} \exp\left\{-\frac{|x - \mu|}{\lambda} + \gamma(x - \mu)\right\} \end{aligned}$$

A.2 Maximizing the log-likelihood function

This subsection deals with the technical details of the maximization of log-likelihoods for the skewed double-exponential model. The likelihood function is

$$L = -\frac{n}{2} \{ \log(V_0) + \log(2 + \gamma^2 V_0) \} \\ - \sum_{i=1}^n |x_i - \mu| \sqrt{2 + \gamma^2 V_0 / V_0} \\ + \gamma \sum_{i=1}^n (x_i - \mu)$$

where m is the number of observations where x_i is greater than μ , l is the number of observations otherwise, and $n = m + l$.

The first-order condition can be derived as follows:

$$\frac{dL}{d\mu} = \sum_{i=1}^n \frac{d|x_i - \mu|}{d\mu} \left(\sqrt{\frac{2 + \gamma^2 V_0}{V_0}} \right)$$

$$+ \gamma \sum_{i=1}^n (-1) \\ = -n\gamma - \sqrt{\frac{2 + \gamma^2 V_0}{V_0}} \\ \times \sum_{i=1}^n \frac{d|x_i - \mu|}{d\mu} = 0$$

$$\frac{dL}{d\gamma} = -\frac{n}{2} \left(\frac{2\gamma V_0}{2 + \gamma^2 V_0} \right) \\ - \sum_{i=1}^n |x_i - \mu| \left(\frac{\gamma}{\sqrt{(2 + \gamma^2 V_0)/V_0}} \right) \\ + \sum_{i=1}^n (x_i - \mu)$$

$$= -\frac{n\gamma V_0}{2 + \gamma^2 V_0} \\ - \sum_{i=1}^n \gamma |x_i - \mu| \sqrt{\frac{V_0}{2 + \gamma^2 V_0}} \\ + \sum_{i=1}^n (x_i - \mu) = 0 \\ \frac{dL}{dV_0} = -\frac{n}{2} \left(\frac{1}{V_0} + \frac{\gamma^2}{2 + \gamma^2 V_0} \right) \\ - \sum_{i=1}^n |x_i - \mu| \\ \times \left(\frac{(\gamma^2 V_0 - 2 - \gamma^2 V_0)/V_0^2}{2\sqrt{(2 + \gamma^2 V_0)/V_0}} \right) \\ = -\frac{n(1 + \gamma^2 V_0)}{V_0(2 + \gamma^2 V_0)} \\ + \sum_{i=1}^n \frac{|x_i - \mu|}{V_0 \sqrt{V_0(2 + \gamma^2 V_0)}} = 0$$

Solving the above first-order conditions gives

$$\mu = \frac{m^2 \sum_{x \in A1} x_i + l^2 \sum_{x \in A2} x_i}{nml} \\ V_0 = \frac{2ml \left(\frac{1}{n} \sum_{i=1}^n x_i - \mu \right)^2}{(l - m)^2} \\ \gamma = \frac{(l - m)^2}{2ml \left(\frac{1}{n} \sum_{i=1}^n x_i - \mu \right)}$$

where $A1 = \{x : x_i - \mu \geq 0\}$ and $A2 = \{x : x_i - \mu < 0\}$.

Given a dataset, we may not be able to find a value of μ to satisfy the above equations. However, by choosing μ to be as close as given by the first-order conditions, we can fit the data well with the skewed double exponential distribution.

Notes

- ¹ See the structural models of Merton (1974), Geske (1977), Leland (1994), Longstaff and Schwartz (1995), and reduced form models of Duffie and Singleton (1999), Jarrow and Turnbull (1995), Das and Tufano (1996), Jarrow *et al.* (1997), Madan and Unal (1999), and Das and Sundaram (2000).
- ² A CDO securitization comprises a pool of bonds (the “collateral”) against which tranches of debt are issued with varying cashflow priority. Tranches vary in credit quality, from AAA to B, depending on subordination level. Seller’s interest is typically maintained as a final equity tranche, carrying the highest risk. CDO collateral usually comprises from a hundred to over a thousand bond issues. Tranche cashflows critically depend on credit events during the life of the CDO, requiring Monte Carlo simulation (see, for example, Duffie and Singleton (1999)) of the joint default process for all issuers in the collateral. An excellent discussion of the motivation for CDOs and the analysis of CDO value is provided in Duffie and Garleanu (2001). For a parsimonious model of bond portfolio allocations with default risk, see Wise and Bhansali (2001).
- ³ As demonstrated by Das *et al.* (2001b), Gersbach and Lippuner (2000), both the individual default probability and default correlations have a significant impact on the value of a credit portfolio.
- ⁴ Extreme value distributions allow for the fact that processes tend to evidence higher correlations when tail values are experienced. This leads to a choice of fatter-tailed distributions, with a reasonable degree of “tail-dependence.” The support of the distribution should also be such that the PD intensity $\lambda_i(t)$ lies in the range $[0, \infty)$.
- ⁵ There is a growing literature on copulas. The following references contain many useful papers on the topic: Embrechts *et al.* (1997), Embrechts *et al.* (1999a,b), Frees and Valdez (1998), Frey and McNeil (2001), Li (1999), Nelsen (1999), Lindskog (2000), Schonbucher and Schubert (2001), Wang (2000).
- ⁶ Durrleman *et al.* (2000) discuss several parametric and non-parametric copula estimation methods.
- ⁷ Lower tail-dependence is symmetrically defined. The coefficient of lower tail-dependence is

$$\lambda_L = \lim_{z \rightarrow 0} \Pr[X_2 < F_2^{-1}(z) | X_1 < F_1^{-1}(z)] \quad (5)$$

If $\lambda_L > 0$, then lower tail-dependence exists.

- ⁸ This specification does permit the hazard rate to populate negative values. While this is not an issue during the estimation phase, the simulation phase is adjusted to truncate

the shock if the negative support is accessed. However, we remark that this occurs in very rare cases, since $\lambda_k(t)$ is the average across all issuers within the rating class, and the averaging drives the probability of negative hazard rates to minuscule levels.

- ⁹ Ang and Chen (2002) found this type of model to be good at capturing the three stated features of asymmetric correlation in equity markets.
- ¹⁰ We found that the estimations of θ_r were very sensitive to the initial values used in the optimization. By fixing these two values, the estimation turned out to be more stable.
- ¹¹ The estimation of Eq. (11) relies on the assumption that the mean around which the individual hazard rate oscillates depends on the initial rating class of the issuer, even though this may change over time. Extending the model to map default probabilities to rating classes is a non-trivial problem, and would complicate the estimation exercise here beyond the scope of this paper. Indeed, this problem in isolation from other estimation issues is complicated enough to warrant separate treatment and has been addressed in a paper by Das *et al.* (2002).
- ¹² This result is consistent with that of Ang and Chen (2002), who undertake a similar exercise with equity returns.
- ¹³ The fact that the Student’s t marginal distribution works best for the jump models may indicate that the jump-model does not capture the dependences as well as the regime-switching model does, since the Student’s t distribution is good at injecting excess kurtosis into the conditional distribution of the marginals.
- ¹⁴ We are grateful to Darrell Duffie for this suggestion.

References

- Abramowitz, M. and Stegun, I. A. (1972). “*The Handbook of Mathematical Functions.*” New York: Dover.
- Ang, A. and Chen, J. (2002). “Asymmetric Correlation of Equity Portfolios.” *Journal of Financial Economics* 63(3), 443–494.
- Bouye, E., Durrleman, V., Nikeghbali, A., Riboulet, G., and Roncalli, T. (2000). “Copulas for Finance: A Reading Guide and some Applications.” Working Paper, Credit Lyonnais, Paris.
- Clayton, D. G. (1978). “A Model for Association in Bivariate Life Tables and its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence.” *Biometrika* 65, 141–151.
- Crouhy, M., Galai, D., and Mark, R. (2000). “A Comparative Analysis of Current Credit Risk Models.” *Journal of Banking and Finance* 24, 59–117.

- Das, S., Fan, R., and Geng, G. (2002). "Bayesian Migration in Credit Ratings based on Probabilities of Default." *Journal of Fixed Income* 12(3), 17–23.
- Das, S., Freed, L., Geng, G., and Kapadia, N. (2001a). "Correlated Default Risk." Working Paper, Santa Clara University.
- Das, S., Fong, G., and Geng, G. (2001b). "The Impact of Correlated Default Risk on Credit Portfolios." *Journal of Fixed Income* 11(3), 9–19.
- Das, S. and Sundaram, R. (2000). "A Discrete-Time Approach to Arbitrage-Free Pricing of Credit Derivatives." *Management Science*, 46(1), 46–62.
- Das, S. and Tufano, P. (1996). "Pricing Credit Sensitive Debt when Interest Rates, Credit Ratings and Credit Spreads are Stochastic." *The Journal of Financial Engineering* 5(2), 161–198.
- Das, S. and Uppal, R. (2000). "Systemic Risk and International Portfolio Choice." Working Paper, London Business School.
- Davis, M. and Violet, L. (1999a). "Infectious Default." Working Paper, Imperial College, London.
- Davis, M. and Violet, L. (1999b). "Modeling Default Correlation in Bond Portfolios." In C. Alexander (ed.) *ICBI Report on Credit Risk*.
- Dowd, K. (1999). "The Extreme Value Approach to VaR: An Introduction." *Financial Engineering News*, Issue 11.
- Duffie, D. J. and Garleanu, N. (2001). "Risk and Valuation of Collateralized Debt Obligations." *Financial Analysts Journal* 57(1), 41–59.
- Duffie, D. J. and Singleton, K. J. (1999). "Simulating Correlated Defaults." Working Paper, Stanford University, Graduate School of Business.
- Durrleman, V., Nikeghbali, A., and Roncalli, T. (2000). "Which Copula is the Right One?" Working Paper, Groupe de Recherche Operationnelle, Credit Lyonnais, France.
- Embrechts, P., Kluppelberg, C., and Mikosch, T. (1997). "Modeling Extremal Events for Insurance and Finance." Berlin: Springer-Verlag.
- Embrechts, P., McNeil, A., and Straumann, D. (1999a). "Correlation and Dependence in Risk Management: Properties and Pitfalls." Working Paper, University of Zurich.
- Embrechts, P., McNeil, A., and Straumann, D. (1999). "Correlation: Pitfalls and Alternatives." Working Paper, Department Mathematik, ETH Zentrum, Zurich.
- Embrechts, P., Lindskog, F., and McNeil, A. (2001). "Modeling Dependence with Copulas and Applications to Risk Management." Working Paper CH-8092, Dept. of Mathematics, ETZH Zurich.
- Frees, E. W. and Valdez, E. A. (1998). "Understanding Relationships using Copulas." *North American Actuarial Journal* 2(1), 1–25.
- Frey, R. and McNeil, A. J. (2001a). "Modeling Dependent Defaults." Working Paper, University of Zurich.
- Frey, R. and McNeil, A. J. (2001b). "Modeling Dependent Defaults: Asset Correlation is not Enough." Working Paper, University of Zurich.
- Genest, C. (1987). "Frank's Family of Bivariate Distributions." *Biometrika* 74, 549–555.
- Genest, C. and Rivest, L. (1993). "Statistical Inference Procedures for Bivariate Archimedean Copulas." *Journal of the American Statistical Association* 88, 1034–1043.
- Geske, R. (1977). "The Valuation of Corporate Liabilities as Compound Options." *Journal of Financial and Quantitative Analysis* 12(4), 541–552.
- Gersbach, H. and Lipponer, A. (2000). "The Correlation Effect." Working Paper, University of Heidelberg.
- Gumbel, E. J. (1960). "Distributions des valeurs extremes en plusieurs dimensions." *Publ. Inst. Statist. Univ. Paris* 9, 171–173.
- Jarrow, R. A. and Turnbull, S. M. (1995). "Pricing Derivatives on Financial Securities Subject to Credit Risk." *Journal of Finance* 50(1), 53–85.
- Jarrow, R. A., Lando, D., and Turnbull, S. M. (1997). "A Markov Model for the Term Structure of Credit Spreads." *Review of Financial Studies* 10, 481–523.
- Jarrow, R. A. and Yu, F. (2000). "Counterparty Risk and the Pricing of Defaultable Securities." Working Paper, Johnson GSM, Cornell University.
- Kijima, M. (2000). "Credit Events and the Valuation of Credit Derivatives of Basket Type." *Review of Derivatives Research* 4, 55–79.
- Lee, A. J. (1993). "Generating Random Binary Deviates having fixed Marginal Distributions and Specified Degree of Association." *The American Statistician* 47, 209–215.
- Leland, H. E. (1994). "Corporate Debt Value, Bond Covenants and Optimal Capital Structure." *Journal of Finance* 49(4), 1213–1252.
- Li, D. X. (1999). "On Default Correlation: A Copula Function Approach." Working Paper 99-07, The RiskMetrics Group, New York.
- Lindskog, F. (2000). "Modeling Dependence with Copulas and Applications to Risk Management." Working Paper, Risklab, ETH Zurich.
- Longin, F. and Solnik, B. (2001). "Extreme Correlation of International Equity Markets." *Journal of Finance* 56, 649–676.

- Longstaff, F. A. and Schwartz, E. S. (1995). "A Simple Approach to Valuing Risky Fixed and Floating Rate Debt." *Journal of Finance* 50(3), 789–819.
- Madan, D. and Unal, H. (1999). "Pricing the Risks of Default." *Review of Derivatives Research* 2(2/3), 121–160.
- Merton, R. (1974). "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *Journal of Finance* 29, 449–470.
- Nelsen, R. B. (1999). "An Introduction to Copulas." New York: Springer-Verlag.
- Sklar, A. (1959). "Functions de repartition a n dimensions et leurs marges." *Publ. Inst. Statist. Univ. Paris* 8, 229–231.
- Sklar, A. (1973). "Random Variables, Joint Distributions, and Copulas." *Kybernetika* 9, 449–460.
- Schonbucher, P. and Schubert, D. (2001). "Copula-Dependent Default Risk in Intensity Models." Working Paper, Dept. of Statistics, Bonn University.
- Wang, S. S. (2000). "Aggregation of Correlated Risk Portfolios: Models and Algorithms." Working Paper, CAS.
- Wise, M. and Bhansali, V. (2001). "Portfolio Allocation to Corporate Bonds with Correlated Defaults." Working Paper No. CALT-68-2365, California Institute of Technology.
- Xiao, J. (2003). "Constructing the Credit Curves: Theoretical and Practical Considerations." Working Paper, RiskMetrics.
- Zhou, C. (2001). "An Analysis of Default Correlations and Multiple Default." *Review of Financial Studies* 14, 555–576.

Keywords: Correlated default; copulas; tail dependence