World Scientific
www.worldscientific.com

EMPIRICAL COPULAS FOR CDO TRANCHE PRICING USING RELATIVE ENTROPY

MICHAEL A. H. DEMPSTER[*], ELENA A. MEDOVA[†] and
SEUNG W. YANG[‡]

Centre for Financial Research
Judge Business School
University of Cambridge
United Kingdom

and

Cambridge Systems Associates Limited
Cambridge, United Kingdom
[*]*mahd2@cam.ac.uk.*
[†]*eam28@cam.ac.uk.*
[‡]*swy21@cam.ac.uk.*

We discuss the general optimization problem of choosing a copula with minimum entropy relative to a specified copula and a computationally intensive procedure to solve its dual. These techniques are applied to constructing an empirical copula for CDO tranche pricing. The empirical copula is chosen to be as close as possible to the industry standard Gaussian copula while ensuring a close fit to market tranche quotes. We find that the empirical copula performs noticeably better than the base correlation approach in pricing non-standard tranches and that the market view of default dependence is influenced by maturity.

Keywords: Portfolio credit risk; CDO; copula; entropy; non-parametric estimation.

1. Introduction

Copula methods for pricing *collateral debt obligations* (CDOs) in general assume some parametric form for the copula of default times and try to obtain the values for model parameters which produce prices that most closely match those of the market. An unsatisfactory aspect of these methods is that they offer little underlying rationale for copula choice. This paper shows how to choose the copula empirically by optimizing its entropy. The strength of the entropic approach is that it provides an information-theoretic rationale for the choice of the copula and also results in excellent fits to data. By minimizing the relative entropy with respect to the industry

---

[*]Corresponding author.

standard Gaussian distribution, we choose the copula that is closest to the standard while ensuring a close fit to market prices.

Our method is similar in spirit to Hull and White [13], who imply from market data an empirical copula in the standard one-factor framework using the criterion of maximal smoothness. We essentially follow the same methodology as theirs but in a more general framework using the criterion of minimum relative entropy. Hull and White's method is limited to calibration to single-maturity data, and is not easily extendable to non-constant hazard rates. The entropic copula approach, however, can be used to calibrate to data across different maturities and naturally accommodates any stochastic hazard rate model. Both these methods promise perfect to near-perfect fits to the data.

The remainder of this introduction discusses credit risk modeling and CDO tranches. Section 2 describes the principle of minimum relative entropy and how to express it numerically and Sec. 3 discusses the maximum entropy copula problem. This is extended to the minimum relative entropy copula problem for CDO tranche pricing in Sec. 4, where computational results based on market data are presented. Section 5 concludes.

1.1. *Correlated intensities in portfolio credit risk modeling*

In *single-name* credit risk modeling, there are two main approaches: the *structural approach* and the *reduced form approach*. The latter has been more popular in pricing applications because it generally offers better fits. A typical example of the reduced form approach is to assume that default occurs when a doubly-stochastic Poisson process (also called a Cox process) first makes a jump.

Extending credit risk modeling to *multiple names* introduces an extra complication. The interdependence between firms in their probability of default is an important aspect that must be taken into account. Some early reduced form approaches attempted to model this dependence by allowing the stochastic intensities of the Cox processes to be correlated and the default events conditioned on the intensities to be independent. Several examples of these models can achieve relatively close fits to market data, for example Mortensen [18] and Graziano and Rogers [11]. As Mortensen [18] conceded, however, the resulting model parameters can be unrealistic because an unnaturally high degree of correlation between the intensities is needed to reproduce the observed market prices. Moreover, Das *et al.* [6] in their empirical study concluded that the level of default dependence that can be realistically introduced by this technique is not sufficient to capture the clustering of defaults that are observed in the market. Why this is the case can be appreciated when we remember that the probabilities of default we are dealing with are very low. To achieve significant clustering of defaults, the default probabilities must be wildly fluctuating at unrealistic levels, as well as being highly correlated.

1.2. *Copulas*

The most popular method for modeling *portfolio credit risk* has been to use copulas. Copulas are used to introduce dependence between default times in a direct way, not indirectly through default intensities. This allows us to reproduce the level of clustering of defaults that we observe in reality. However, as mentioned previously, the choice of copula is rather arbitrary, motivated by two main criteria: the quality of final fit to the data, and computational tractability. As we shall see, choosing an empirical copula using the entropic approach gives us an underlying rationale for this choice.

1.3. *CDO tranche pricing*

A CDO is a derivative structure which provides protection against the loss on a portfolio of defaultable assets. The seller of protection on a tranche of this portfolio receives regular premium payments. In return, he must pay the buyer of protection any losses on that tranche that are incurred through defaults. Each tranche covers only a portion of the total potential losses of the portfolio.

To illustrate, consider a CDO portfolio of n *names*, each with *unit nominal* amount and with *maturity T*. Then denoting $\boldsymbol{\tau}_j$ as the *time of default of name j* the amount lost on the portfolio at time t is

$$L(t, \boldsymbol{\tau}) = (1 - R) \sum_{j=1}^{n} 1_{\{\boldsymbol{\tau}_j < t\}}, \tag{1.1}$$

where R is the constant *recovery rate* and $\boldsymbol{\tau} = (\boldsymbol{\tau}_1, \dots, \boldsymbol{\tau}_n)$ is the vector of *default times*.[1] Now consider a *tranche* of this CDO with *attachment point α* and *detachment point β*. Then the *loss* on this tranche at time t will be the non-decreasing function

$$M(t, \boldsymbol{\tau}) = \begin{cases} 0 & \text{if } L(t, \boldsymbol{\tau}) \leq \alpha, \\ L(t, \boldsymbol{\tau}) - \alpha & \text{if } \alpha < L(t, \boldsymbol{\tau}) \leq \beta, \\ \beta - \alpha & \text{if } L(t, \boldsymbol{\tau}) > \beta. \end{cases}$$
$$= \beta - \alpha - (\beta - L(t, \boldsymbol{\tau}))^+ + (\alpha - L(t, \boldsymbol{\tau}))^+,$$

where $(x)^+ := \max\{x, 0\}$. Thus the loss on an (α, β) tranche can be written in terms of a *put spread* as shown in Fig. 1 below. Then the *default leg* — the present value of the (random) amount that the *seller* of protection needs to pay — can be written as

$$D(\boldsymbol{\tau}, T) = \sum_{j=1}^{n} e^{-r\boldsymbol{\tau}_j} 1_{\{\boldsymbol{\tau}_j \leq T\}} \left(M(\boldsymbol{\tau}_j, \boldsymbol{\tau}) - M(\boldsymbol{\tau}_j^-, \boldsymbol{\tau}) \right) \tag{1.2}$$

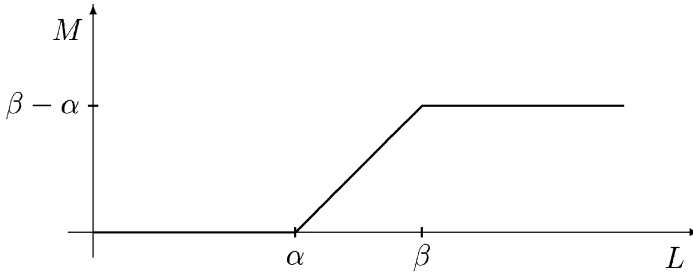[1]We use boldface throughout to denote random entities.

Fig. 1. Loss on a (α, β) tranche vs portfolio loss.

and the *premium leg* — the present value of the (random) amount that the seller of protection receives in return — can be written as

$$P(\boldsymbol{\tau}, T) = s \sum_{k=1}^{P} e^{-rt_k} \left(\beta - \alpha - M(t_k, \boldsymbol{\tau}) \right), \tag{1.3}$$

where s is the premium rate, and where $t_k \leq T, k = 1, \ldots, P$, are the premium payment dates.[2]

To price a CDO tranche, we wish to work out the fair *premium rate* $\bar{s}$. Like any other swap, this is the value that makes the expected payoffs of the default and premium legs equal. Therefore the fair premium rate is the value $\bar{s}$ which satisfies

$$E[D(\boldsymbol{\tau}, T)] - \bar{s}E[P(\boldsymbol{\tau}, T)] = 0. \tag{1.4}$$

It should be clear that it is the "tranching" feature which makes the fair premium rate $\bar{s}$ depend on the *default dependence*. If we consider a "tranche" that spans the whole portfolio, i.e., a $(0, L_{\max})$ tranche, then we would have $M \equiv L$; since L is linear in each of the default times, the expected payoffs $E[D(\boldsymbol{\tau}, T)]$ and $E[P(\boldsymbol{\tau}, T)]$ only depend on the mean of $\boldsymbol{\tau}$ and not its higher moments, and hence the fair premium rate $\bar{s}$ does not depend on the default dependence. This is not the case if the tranche only spans a subset of the overall portfolio loss.

2. Minimum Relative Entropy

The principle of minimum relative entropy is closely related to the principle of maximum entropy and it is instructive to consider the latter first.

2.1. *Principle of maximum entropy*

The *principle of maximum entropy* (MaxEnt) is a method of obtaining a unique probability distribution for a random variable from a given set of data assumed to be generated by it. The principle was first formulated by Jaynes [15] and is used in a wide variety of applied sciences.

---

[2]We ignore accrued payments here for expositional simplicity but include them in our numerical studies.

To illustrate a typical problem in finance, suppose we have a finite set of vanilla European option prices on a stock. The price of each option is a function of the risk-neutral density of the stock price at maturity. The aim of MaxEnt in this example is to infer from the given set of option prices the risk-neutral density of the stock price.

However, such a problem is generally highly under-determined because we have many fewer option prices than possible stock prices. But a unique density can be obtained if we optimize some objective function that depends on the density, while satisfying the observed market prices. Various forms for the objective function have been proposed, such as Fisher information [9, 12], maximal smoothness [13, 14] and entropy [4].

Since entropy is a measure of uncertainty, inferring the probability distribution of a random variable by maximizing entropy is optimal in the sense that we only take into account information that is given and do not assume anything else about the distribution. We choose a distribution which is consistent with the given information but otherwise has maximum uncertainty. MaxEnt applied to finance is thus related to the concept of *market efficiency* in the sense that prices fully reflect all available information in the market.

The MaxEnt principle is a *non-parametric* method of estimating a probability distribution. In *parametric* estimation the focus is on obtaining the best estimator $\hat{\theta}$ of a given parametric family $f(\cdot|\theta)$ of densities. This process involves two steps: model *specification* and model *estimation*. In non-parametric estimation the focus is on obtaining a good estimate of f *directly* from the data, which eliminates the need for model specification.

The MaxEnt principle is thus well-suited to the estimation of copulas in portfolio credit risk modelling. As discussed earlier, the choice of copula for CDO tranche pricing in much of the literature is rather arbitrary, motivated by quality of fit. These methods are parametric in nature. By using a non-parametric approach such as MaxEnt the problem of arbitrarily choosing a copula is obviated.

To set notation, suppose we observe the data set $\{(\tilde{a}_1, \bar{a}_1), \ldots, (\tilde{a}_m, \bar{a}_m)\}$ with $\tilde{a}_i, \bar{a}_i \in \mathbb{R}$ for each $i = 1, \ldots, m$. We know from the problem at hand that $E[a_i(\mathbf{x}, \bar{a}_i)] = \tilde{a}_i$ for each i, where $\mathbf{x}$ is a random vector taking values in some domain $D \subseteq \mathbb{R}^n$.

Letting f be the density for $\mathbf{x}$ (or probability mass function if discrete) we wish to maximize its *differential entropy*, i.e., solve the problem

$$\sup_{f \in L_1^+(D)} -\int_D f(x) \log f(x) dx, \tag{2.1}$$

subject to the data constraints

$$\int_D a_i(x, \bar{a}_i) f(x) dx = \tilde{a}_i \quad i = 1, \ldots, m, \tag{2.2}$$

where $a_i(\cdot, \bar{a}_i)$ is a piecewise continuous function on D and $L_1^+(D)$ denotes the non-negative cone of the space of real-valued integrable functions on D with the usual integral norm.

2.2. *Solution to the MaxEnt problem*

The Lagrangian for this problem can be easily maximized by elementary calculus of variations.[3,4] After normalizing f to make it a probability density the solution is of the form

$$\hat{f}(x, \lambda) = \frac{1}{Z(\lambda)} \exp \left\{ \sum_{i=1}^{m} \lambda_i a_i(x, \bar{a}_i) \right\}, \tag{2.3}$$

where $\lambda = (\lambda_1, \ldots, \lambda_m)$ and $Z(\lambda) := \int \exp\{\sum_i \lambda_i a_i(x, \bar{a}_i)\}dx$ is the normalizing constant. The optimal f can then found by solving for the unique values λ_i that satisfy the constraints (2.2).

However, instead of solving the m simultaneous equations (2.2) with $f := \hat{f}$ to determine the lambdas, we can more elegantly solve the dual problem. Consider the dual function

$$\mathcal{L}^*(\lambda) = - \int_D \hat{f}(x) \log \hat{f}(x)dx + \int_D \lambda' \left(A(x) - \tilde{a} \right) \hat{f}(x, \lambda)dx \tag{2.4}$$

$$= \log Z(\lambda) - \lambda'\tilde{a}, \tag{2.5}$$

where $A(x) := (a_1(x, \bar{a}_1), \ldots, a_m(x, \bar{a}_m))'$ and $\tilde{a} = (\tilde{a}_1, \ldots, \tilde{a}_m)'$ and the prime symbol denotes transpose. The *dual problem* of the MaxEnt problem given by (2.1) and (2.2) is

$$\inf_{\lambda \in \mathbb{R}^m} \mathcal{L}^*(\lambda).$$

This is a much easier optimization problem than the original primal problem because it is finite dimensional and unconstrained, and we know that the dual function is always convex. Moreover, we can easily show that it is *strictly convex* if the functions a_i are linearly independent.

Proposition 2.1. *The dual function (2.4) is convex, and it is strictly convex if and only if the functions a_i are linearly independent.*

Proof. Take $\bar{\lambda}$ and $\tilde{\lambda}$ in $\mathbb{R}^m$ and set $\lambda := s\bar{\lambda} + (1-s)\tilde{\lambda}, s \in [0, 1]$. Then by Hölder's inequality

$$\mathcal{L}^*(\lambda) = \log \int \exp \left\{ s\bar{\lambda}'A(x) + (1-s)\tilde{\lambda}'A(x) \right\} dx - s\bar{\lambda}'\bar{a} - (1-s)\tilde{\lambda}'\bar{a}$$

$$\leq s \left(\log \int e^{\bar{\lambda}'A(x)}dx - \bar{\lambda}'\bar{a} \right) + (1-s) \left(\log \int e^{\tilde{\lambda}'A(x)}dx - \tilde{\lambda}'\bar{a} \right).$$

If the a_i's are not linearly independent then we can find a $\bar{\lambda}$ and $\tilde{\lambda}$ with $\bar{\lambda} \neq \tilde{\lambda}$ such that $\bar{\lambda}'A \equiv \tilde{\lambda}'A$, whence Hölder's inequality becomes equality. But if the a_i's are linearly independent, then seeing Hölder's inequality as an application of Jensen's

---

[3] To apply the Euler–Lagrange equation, i.e., equate the first variation to zero, we need to assume that there exists a feasible f with finite entropy and that $f > 0$ almost everywhere.
[4] Note that the Lagrangian is concave in f.

inequality for the strictly concave function given by $f(x) := x^s$ for $s \in (0, 1)$, the result follows.

Alternatively we could show that the Hessian is the covariance matrix of the a_i's, and thus the dual function is strictly convex if the a_i's are linearly independent. $\qquad\square$

The minimization can be solved numerically using any gradient-based optimization method such as the BFGS quasi-Newton algorithm.[5] We just need the gradient vector of the dual, whose ith element is given by

$$\frac{\partial \mathcal{L}^*(\lambda)}{\partial \lambda_i} = \int_D a_i(x, \bar{a}_i) \hat{f}(x, \lambda) dx - \tilde{a}_i.$$

We can differentiate the dual (2.4) under the integral sign since the integrand is in $L_1(D)$ for each fixed λ, and is differentiable with respect to λ for almost all x with bounded derivative for all bounded λ.

Notice that the gradient corresponds to the constraints (2.2) and that the constraints will be satisfied when the gradient vector is zero.

In addition to the assumption that the feasible set is not empty, we also impose the assumption mentioned previously in footnote 3 that there must exist a feasible f with $f > 0$ almost everywhere which has finite entropy.

2.3. *Regularization*

The Slater constraint qualification that we have just mentioned is difficult to check in general. In particular, it is difficult to determine whether the observed data set $\{(\tilde{a}_1, \bar{a}_1), \ldots, (\tilde{a}_m, \bar{a}_m)\}$ is consistent.[6] A related problem is that in real applications there may be measurement errors in the observed data — not only are these errors a problem in themselves, but they may also cause inconsistencies in the data which render the feasible set empty.

Both these problems can be overcome if we consider a penalized version of the MaxEnt problem

$$\sup_{f \in L_1^+(D)} -\int_D f(x) \log f(x) dx - \frac{1}{2\theta} \sum_{i=1}^m w_i^2 \left[\int_D a_i(x, \tilde{a}_i) f(x) dx - \bar{a}_i \right]^2 \qquad (2.6)$$

for some positive θ and w_i. Here θ plays the role of "temperature" — the lower we set θ, the smaller the errors will be. The w_i act as weights to emphasize the importance of a particular constraint — the higher we set w_i, the smaller the error

---

[5]The BFGS algorithm was independently developed by Broyden [3], Fletcher [8], Goldfarb [10] and Shanno [19].
[6]Borwein *et al.* [2] have however developed an easy-to-check test for this constraint qualification in the standard MaxEnt problem for call options.

for the ith constraint will be. Although this problem is unconstrained, as discussed in Decarreau *et al.* [7], we can reformulate (2.6) in the form of the original problem as

$$\sup_{f \in L_1^+(D), \epsilon \in \mathbb{R}^m} - \int_D f(x) \log f(x) dx - \frac{1}{2\theta} \|\epsilon\|^2$$

subject to

$$\int_D a_i(x, \bar{a}_i) f(x) dx - \tilde{a}_i = \frac{\epsilon_i}{w_i}, \quad i = 1, \dots, m,$$

where $\| \cdot \|$ denotes the Euclidean norm on $\mathbb{R}^m$. Then one can easily show that the dual function of this penalized problem is given by

$$\mathcal{L}^*(\lambda) = \log Z(\lambda) - \lambda' \tilde{a} + \frac{1}{2}\theta \|\tilde{\lambda}\|^2,$$

where $\tilde{\lambda} = (\lambda_1/w_1, \dots, \lambda_m/w_m)'$, and the ith element of its gradient vector is given by

$$\frac{\partial \mathcal{L}^*(\lambda)}{\partial \lambda_i} = \int_D a_i(x, \bar{a}_i) \hat{f}(x, \lambda) dx - \tilde{a}_i + \theta \frac{\lambda_i}{w_i^2}.$$

Thus we can still use the dual approach as discussed above to solve this problem.

The above technique is known as *penalization*. Another approach, called *relaxation*, is to solve the problem

$$\sup_{f \in L_1(D)} - \int_D f(x) \log f(x) dx$$

subject to

$$\left| \int_D a_i(x, \bar{a}_i) f(x) dx - \tilde{a}_i \right| \le \epsilon, \quad i = 1, \dots, m$$

for some $\epsilon > 0$. The problem with relaxation, as opposed to penalization, is that the issue of consistency remains — the feasible set may still be empty if ϵ is too small. We will use penalization for our application.

2.4. *Principle of minimum relative entropy*

Since the distribution with the greatest entropy is the uniform distribution,[7] when we apply the principle of maximum entropy we are effectively choosing the distribution that is "closest" to the uniform distribution while satisfying the data constraints. But we could, if we wish, choose a distribution other than the uniform distribution. To do this we use the concept of *relative entropy*.

---

[7]Among distributions with bounded support.

Relative entropy is a measure of "distance" of one probability distribution to another. For absolutely continuous probability distributions F and G it is defined by[8]

$$L(F|G) = \int_D f(x) \log \frac{f(x)}{g(x)} dx,$$

where f and g are the densities of F and G respectively. Minimizing relative entropy will be useful if we have a *prior belief* about what the empirical distribution might or should be.

The form of the optimum solution to the *minimum relative entropy* (MinRelEnt) problem is almost identical to that of the MaxEnt problem and is given by

$$\hat{f}(x, \lambda) = \frac{1}{Z(\lambda)} g(x) \exp \left\{ \sum_{i=1}^{m} \lambda_i a_i(x, \bar{a}_i) \right\}, \qquad (2.7)$$

where $Z(\lambda) := \int g(x) \exp\{\sum_i \lambda_i a_i(x, \bar{a}_i)\} dx$. The optimal f can again be found by solving for the unique values λ_i that satisfy the constraints (2.2), possibly using the dual approach as discussed above.

We need to note the regularity conditions required on the prior g. To apply the Euler–Lagrange equation, we must (corresponding to footnote 3) make the assumption that there is a feasible f with finite relative entropy which is equivalent to g (i.e., their supports agree almost everywhere). The latter is difficult to check but we can ensure that an optimal solution has finite relative entropy by requiring $\int_D |\hat{f}(x) \log \hat{f}(x)/g(x)| dx < \infty$. The actual conditions on g therefore depend on the functions a_i and the domain D.

3. The MaxEnt Copula Problem

If we wish to apply the MaxEnt principle to a problem involving a copula, then in addition to the data constraints (2.2) above we must also require constraints on the marginals of f.

A *copula* is a joint distribution function on the unit hypercube $[0,1]^n$ with marginals that are uniformly distributed. So in addition to the data constraints, we need the *marginal constraints*

$$\int_0^p \int_{[0,1]^{n-1}} f(x) dx_{\neg j} dx_j = p \quad \forall p \in [0,1] \quad j = 1, \ldots, n. \qquad (3.1)$$

We therefore have an infinite dimensional constraint space, unlike the finite dimensional case defined by (2.2).

---

[8]Relative entropy is always non-negative and is equal to zero if and only if $f \equiv g$.

3.1. *Discrete approximation*

One technique to deal with the infinite dimensional constraints is to take only a finite number $p_1, \ldots, p_N \in [0, 1]$ of discrete points and require that the above marginal constraints hold only for this set of points, but not for all $p \in [0, 1]$. Thus we would have instead of (3.1)

$$\int_{p_{k-1}}^{p_k} \int_{[0,1]^{n-1}} f(x) dx_{\neg j} dx_j = \Delta p_k \quad j = 1, \ldots, n, \quad k = 1, \ldots, N,$$

where $\Delta p_k := p_k - p_{k-1}$. This method may be satisfactory for many applications. For example in pricing CDO tranches, if we take each p_k to correspond to quarterly time steps or even annual time steps and the longest maturity of the tranches is 10 years, then we need only take $N = 40$ or 10, respectively, because there is no need to go beyond the longest maturity in the data set. Furthermore, if we assume a homogeneous portfolio where each firm has the same marginal default time distribution, then we know that the Lagrange multipliers for the p_k's for each dimension will be the same. There will thus only be an additional 10 to 40 more constraints since we do not need a separate set of Lagrange multipliers for each dimension.

3.2. *The MaxEnt copula problem*

There are two alternative ways of formulating the MaxEnt copula problem.

For a copula with density c the MaxEnt problem is to maximize its entropy $-\int_{[0,1]^n} c(u) \log c(u) du$, subject to the data constraints being satisfied and the marginals of c being uniformly distributed. By a simple change of variables, we can show that this problem is equivalent to minimizing

$$\int_D f(x) \log \frac{f(x)}{f_1(x_1) f_2(x_2)} dx,$$

where f is the joint density function with support D, and f_1 and f_2 are the marginal density functions of the relevant problem (assuming here just two dimensions for simplicity). Thus we can see that maximizing the entropy of a copula is equivalent to mimizing the entropy of the corresponding joint distribution to be as close as possible to the *independent case*, i.e., $f(x) = f_1(x_1) f_2(x_2)$.

The other way to formulate the MaxEnt copula problem is to start from the joint distribution. Thus we maximize $-\int_D f(x) \log f(x) dx$ subject to the data constraints being satisfied and f having marginals f_1 and f_2. Expressing this in terms of the copula, the problem becomes equivalent to maximizing

$$-\int_{[0,1]^2} c(u) \log \{c(u) f_1(x_1) f_2(x_2)\} du,$$

where $x_i = F_i^{-1}(u_i)$ for $i = 1, 2$ and the F_i are the marginal cdfs.

It would seem that both ways of formulating the problem are equally valid, although the former allows for an easier interpretation.

4. Application to CDO Tranche Pricing

We now apply the principle of minimum relative entropy to CDO tranche pricing. There are several things to consider in specifying this problem.

4.1. *The empirical copula problem for CDO pricing*

We choose to work with the copula of default times rather than with the joint distribution. Writing the CDO tranche pricing equation (1.4) in integral form we have

$$\int_{[0,\infty)^n} D(\tau,T)f(\tau)d\tau - \bar{s}\int_{[0,\infty)^n} P(\tau,T)f(\tau)d\tau = 0. \tag{4.1}$$

Rewriting (4.1) using Sklar's lemma $(f(\tau) = c(u)\Pi f_i(\tau_i))$ yields

$$\int_{[0,1]^n} D\left(F^{-1}(u),T\right)c(u)du - \bar{s}\int_{[0,1]^n} P\left(F^{-1}(u),T\right)c(u)du = 0, \tag{4.2}$$

where in terms of the marginal default times $(F^{-1}(u) := (F_1^{-1}(u_1),\ldots,F_n^{-1}(u_n)) = (\tau_1,\ldots,\tau_n) = \tau)$ and $c(u)$ is the density of the copula of default times (noting that there is cancellation of the Jacobian factor).

Next we must determine the form of the functions a_i in terms of the argument u of the copula density c. Referring back the equations for the default and premium payoffs (1.2), (1.3) and using (4.2) we obtain[9]

$$a_i(\mathbf{u},\bar{s}_i) = \sum_{j=1}^{n} e^{-r\boldsymbol{\tau}_j} 1_{\{\boldsymbol{\tau}_j \leq T\}}\left(M(\boldsymbol{\tau}_j,\boldsymbol{\tau}) - M(\boldsymbol{\tau}_j^{-},\boldsymbol{\tau})\right)$$

$$- \bar{s}_i \sum_{k=1}^{P} e^{-rt_k}(\alpha - \beta - M(t_k,\boldsymbol{\tau})), \tag{4.3}$$

where $\boldsymbol{\tau}_j = F_j^{-1}(\mathbf{u}_j)$. As for the observed data values $\tilde{s}_i$ (1.4) implies that they are always zero except for the equity tranche.[10]

We can now specify the *empirical copula CDO pricing problem*. For computational efficiency we assume a homogenous portfolio. To account for bid-ask spreads and also for inaccuracy from Monte Carlo integration (discussed below) we solve the penalized minimum relative entropy problem as outlined in Sec. 2.3. Thus the problem is

$$\sup_{c\in L^1[0,1]^n, \epsilon\in\mathbb{R}^{m+N}} -\int_{[0,1]^n} c(u)\log\frac{c(u)}{\pi(u)}du - \frac{1}{2\theta}|\epsilon|^2$$

---

[9]We have not mentioned accrued payments for notational simplicity, but they have been included in the numerical studies.

[10]The quote for the equity tranche on the iTraxx and CDX indices represents the upfront payment that must be made, where the running premium is set at $\bar{a}_i = 500$ bps.

subject to the data constraints given by

$$\int_{[0,1]^n} a_i(u, \bar{s}_i) c(u) du - \tilde{s}_i = \frac{\epsilon_i}{w_i} \quad i = 1, \dots, m$$

and the marginal constraints given by

$$\int_{p_{k-1}}^{p_k} \int_{[0,1]^{n-1}} c(u) du_{\neg j} du_j - \Delta p_k = \frac{\epsilon_k}{w_k} \quad k = 1, \dots, N,$$

for each $j = 1, \dots, n$, where $0 = p_0 < \cdots < p_N < 1$ and $\Delta p_k := p_k - p_{k-1}$.[11] Here π represents some *prior* copula for the specified marginal default time distributions. This problem has a unique solution given by

$$\hat{c}(u) = \frac{1}{Z(\lambda)} \pi(u) \exp\left\{ \sum_{i=1}^{m} \lambda_i a_i(u, \bar{s}_i) + n \sum_{k=1}^{N} \lambda_k 1_{\{p_k \le u < p_{k+1}\}} \right\}, \qquad (4.4)$$

where $\lambda \in \mathbb{R}^{m+N}$ with normalizing constant

$$Z(\lambda) = \int_{[0,1]^n} \pi(u) \exp\left\{ \sum_{i=1}^{m} \lambda_i a_i(u, \bar{s}_i) + n \sum_{k=1}^{N} \lambda_k 1_{\{p_k \le u < p_{k+1}\}} \right\} du.$$

The *dual function* $\mathcal{L}^*$ of this problem is given by

$$\mathcal{L}^*(\lambda) = \log Z(\lambda) - \sum_{i=0}^{n} \lambda_i \tilde{s}_i - n \sum_{k=1}^{N} \lambda_k \Delta p_k + \frac{1}{2} \theta |\tilde{\lambda}|^2,$$

where $\tilde{\lambda} = (\lambda_1/w_1, \dots, \lambda_{m+N}/w_{m+N})$, and again is strictly convex with gradient

$$\frac{\partial \mathcal{L}^*(\lambda)}{\partial \lambda_i} = \int_{[0,1]^n} a_i(u, \bar{s}_i) \hat{c}(u) du - \tilde{s}_i + \theta \frac{\lambda_i}{w_i^2}$$

$$\frac{\partial \mathcal{L}^*(\lambda)}{\partial \lambda_k} = n \int_{p_{k-1}}^{p_k} \int_{[0,1]^{n-1}} \hat{c}(u_{\neg j}, u_j) du_{\neg j} - n \Delta p_k + \theta \frac{\lambda_k}{w_k^2}.$$

We must impose the regularity condition on the prior copula π. As stated in Sec. 2.4 we require $\int |\hat{c}(u) \log \hat{c}(u)/\pi(u)| du < \infty$. If we set $\bar{c} := \hat{c}/\pi$, then we need

$$\int |\pi(u) \bar{c}(u) \log \bar{c}(u)| du < \infty. \qquad (4.5)$$

Notice from (4.3) the functions a_i have bounded range. As each $u_j \to 0$ we have $a_i \to 125(1 - R)$. This corresponds to extreme the case where every firm has defaulted instantaneously at time $t = 0$. In the opposite extreme as each $u_j \to 1$ we have $a_i \to -\bar{s}_i PL_{\max}$ which corresponds to the case when no firms default. As a_i's are decreasing functions they are bounded. It follows that the range of $\bar{c}$ is $[b, B]$

---

[11] Note that we do not need to partition all of $[0, 1]$, as mentioned previously in Sec. 3.1. We can take p_N corresponding to the longest maturity in the data set.

for some $b > 0$ and $B < \infty$. Therefore

$$\int_{[0,1]^n} |\pi(u)\bar{c}(u) \log \bar{c}(u)| du \leq \int_{[0,1]^n} |\pi(u)| du \cdot \text{ess sup}_u |\bar{c}(u) \log \bar{c}(u)| < \infty,$$

so we can see that (4.5) is satisfied for *any* prior copula π.

We have assumed that the portfolio is homogeneous, so that each firm has the same marginal default time distribution. Our experiments and also those of Hull and White [13] show that the value of a CDO tranche is not particularly sensitive to whether the portfolio is homogenous or heterogenous — it is the *average* of the default probabilities of all firms that principally determines the value of a CDO tranche.

4.2. *Numerical issues*

An important issue to consider is how to use the resulting empirical copula to compute CDO tranche prices. A typical CDO application would be to calibrate the copula to the iTraxx or CDX tranche quotes. The iTraxx and CDX indices both refer to a portfolio of 125 names. Therefore to compute the fair premium $\bar{s}$ using equation (4.2) requires computing $n = 125$ dimensional integrals. This forces us to use Monte Carlo methods.

With just simple Monte Carlo integration computing CDO prices with the Min-RelEnt copula is very slow to converge. Markov Chain Monte Carlo methods have been tried and converge much faster in some cases but are not robust in general.[12] However, the presence of the factor $\pi(x)$ in (4.4) allows us to use *importance sampling* if the prior is chosen to be a copula from which we can easily simulate. This is effective in speeding up convergence.

We also used an additional importance sampling technique proposed by Joshi [16]. We choose the prior π to be the Gaussian copula and generate random variates from it using the well-known one factor method. We then shift the mean of the common factor by some amount μ to simulate more default times that will affect the payoffs of the tranches and multiply the resulting integrand by the likelihood ratio $\exp\{0.5\mu^2 - \mu X\}$, where X is the realization of the common factor.

4.3. *Calibration to simulated CDO prices*

We conducted two sets of tests. The first set of tests involved generating a set of simulated market CDO quotes from a given copula. After calibrating the minimum relative entropy copula to a subset of these quotes — which we call the training set — we priced the remaining out-of-sample tranches with it and compared them to the known true prices to see how well the minimum relative entropy copula can learn about the true underying copula.

---

[12]Specifically, the Metropolis random walk algorithm [17].

The following assumptions were used.

Number of firms	125
Risk-free rate	0.05
Hazard rate for each firm	0.005
Recovery rate	0.4
Premium payments per year	4

The prior for the minimum entropy copula was chosen to be the Gaussian copula with correlation 0.4.

We simulate the market CDO tranche quotes from the *stochastic correlation copula*. This copula is simply the Gaussian copula with random correlation values and is a good candidate because it is one of the copulas that fits market prices relatively well, as well as being easy to simulate from [5]. We will use the discrete distribution $\rho = (0.066, 0.2, 0.8)$ with probabilities $p = (0.66, 0.1, 0.24)$ for the random correlation parameter.[13]

The training set contains tranches of different maturities and all maturities are calibrated simultaneously. The training set values and the calibration results are given in Table 1. Note that the equity (0,3) tranche premia are not in basis points but are expressed as a percentage of the nominal to be paid upfront. As we can see, the "calibration" is very good.

The next step is to see how well the empirical copula can "interpolate" across tranche threshold levels. The out-of-sample tranche pricing results are shown in

Table 1. Calibration to simulated data.

Thresholds (%)	Maturity (yrs)	True premium (bps)	MinRel premium (bps)	Absolute error (bps)
0–3	5	14.7	14.7	0.0
3–6	5	99.2	99.6	0.4
6–9	5	32.9	33.3	0.5
9–12	5	21.8	22.1	0.3
12–22	5	14.0	13.9	0.1
0–3	7	18.2	18.3	0.1
3–6	7	136.2	136.3	0.1
6–9	7	39.7	39.9	0.2
9–12	7	23.3	23.5	0.2
12–22	7	14.6	14.6	0.0
0–3	10	21.3	21.4	0.1
3–6	10	185.1	185.6	0.5
6–9	10	53.9	54.0	0.1
9–12	10	26.7	26.7	0.0
12–22	10	15.4	15.3	0.1
Total error				2.7

[13]These are the values used in Burtschell *et al.* [5]. There is nothing inherently special about these values other than that they produce reasonable prices.

Table 2. Prices of non-standard tranches using the empirical copula.

Thresholds (%)	Maturity (yrs)	True premium (bps)	MinRel premium (bps)	Absolute error (bps)
1.5–4.5	5	271.1	276.7	5.6
4.5–7.5	5	49.6	50.3	0.7
7.5–10.5	5	25.9	26.2	0.3
10.5–17.0	5	17.2	17.0	0.2
1.5–4.5	7	331.6	336.9	5.3
4.5–7.5	7	65.8	66.1	0.3
7.5–10.5	7	28.7	29.5	0.8
10.5–17.0	7	17.8	18.2	0.4
1.5–4.5	10	389.5	396.9	7.4
4.5–7.5	10	94.1	92.9	1.2
7.5–10.5	10	35.4	34.8	0.6
10.5–17.0	10	19.0	19.0	0.0
Total error				22.8

Table 2. Although it has some trouble pricing the (1.5, 4.5) tranche accurately, overall we can see that the empirical copula performs well out-of-sample. For comparison, we also price these non-standard tranches using the industry standard base correlation approach [1]. The results are shown in Table 3. We can see that although the base correlation method is relatively accurate for the senior mezzanine and super senior tranches, it performs much worse than the MinRelEnt copula for the (1.5, 4.5) tranche. This is likely because the loss distribution associated with the true copula has a high peak in this region (see the discussion below). Overall, the MinRelEnt copula performs significantly better than the base correlation approach.

We also tested to see how well the empirical copula can "interpolate"/ "extrapolate" across maturities. The results are shown in Table 4. Again we can see excellent

Table 3. Prices of non-standard tranches using base correlation.

Thresholds (%)	Maturity (yrs)	True premium (bps)	BaseCorr premium (bps)	Absolute error (bps)
1.5–4.5	5	271.1	255.8	15.3
4.5–7.5	5	49.6	51.1	1.5
7.5–10.5	5	25.9	23.6	2.3
10.5–17.0	5	17.2	18.2	1.0
1.5–4.5	7	331.6	307.2	24.4
4.5–7.5	7	65.8	70.4	4.6
7.5–10.5	7	28.7	28.1	0.6
10.5–17.0	7	17.8	19.1	1.3
1.5–4.5	10	389.5	363.6	25.9
4.5–7.5	10	94.1	99.3	4.8
7.5–10.5	10	35.4	37.2	1.8
10.5–17.0	10	19.0	20.5	1.5
Total error				85.0

Table 4. Prices of tranches with non-standard maturities using the empirical copula.

Thresholds (%)	Maturity (yrs)	True premium (bps)	MinRel premium (bps)	Absolute error (bps)
0–3	3	9.7	9.2	0.5
3–6	3	64.8	70.0	5.2
6–9	3	28.7	34.4	5.7
9–12	3	20.6	25.5	4.9
12–22	3	13.2	18.5	5.3
0–3	4	12.4	12.2	0.2
3–6	4	81.2	82.0	0.8
6–9	4	30.6	32.2	1.6
9–12	4	21.3	22.4	1.1
12–22	4	13.7	15.1	1.4
0–3	6	16.6	16.6	0.0
3–6	6	117.5	117.1	0.4
6–9	6	35.8	35.8	0.0
9–12	6	22.3	22.0	0.3
12–22	6	14.1	14.1	0.0
0–3	8	19.5	19.4	0.1
3–6	8	153.5	151.8	1.7
6–9	8	43.8	43.6	0.2
9–12	8	24.1	24.7	0.6
12–22	8	14.8	14.6	0.2
0–3	9	20.5	20.5	0.0
3–6	9	169.8	170.0	0.2
6–9	9	48.4	49.0	0.6
9–12	9	25.2	26.1	0.9
12–22	9	15.0	15.2	0.2

performance of the empirical copula in interpolating across non-standard maturities. It even performs reasonably well extrapolating to maturities less than five years, although we can see the accuracy does start to deteriorate for the three-year maturity.

4.4. *Calibration to market CDO prices*

In the second set of tests, we examine whether or not the MinRelEnt copula can be accurately calibrated to market data. We first attempt to calibrate to each maturity separately and then to all maturities simulaneously to see how much accuracy is lost. If the calibration across all maturities is much poorer than the calibration for a single maturity, then this would suggest that the market does not price the dependency as static. We also calibrate to the 5 and 10 year maturities, and see how well these calibrations can match the 7 year prices out-of-sample.

The recovery rate is assumed to be the same as before, but the default rate curve is now determined from the market index quotes and is assumed to be piecewise constant between maturities. We also replace the constant risk-free rate used in the simulation experiment by the market swap curve.

Table 5. Calibration to iTraxx, each maturity separately.

Thresholds (%)	Maturity (yrs)	Bid-ask (bps)	Market premium (bps)	MinRel premium (bps)	Absolute error (bps)
0–3	5	0.5	23.3	24.5	1.2
3–6	5	2.0	68.0	69.6	1.6
6–9	5	2.0	19.0	18.8	0.2
9–12	5	3.0	9.5	8.9	0.6
12–22	5	0.5	4.3	4.1	0.2
0–3	7	0.5	43.1	43.9	0.8
3–6	7	4.0	202.0	203.4	1.4
6–9	7	3.0	46.5	45.5	1.0
9–12	7	2.0	24.0	22.9	1.1
12–22	7	2.0	9.0	7.9	1.1
0–3	10	0.5	54.3	54.3	0.0
3–6	10	10.0	580.0	580.4	0.4
6–9	10	6.0	117.0	117.1	0.1
9–12	10	3.0	51.5	51.6	0.1
12–22	10	2.0	20.0	20.1	0.1

Table 5 shows the calibration results for the iTraxx tranche quotes for 4 April 2006.[14] Although not quite as good as the results for the simulated data, we can see that the calibration errors lie mainly within the bid-ask spreads. Interestingly, it is again for the junior mezzanine (3,6) tranche where the MinRelEnt copula is least accurate. Similar performance was achieved for CDX data.

Next we attempted to calibrate to all maturities simultaneously, the results of which are shown in Table 6. The accuracy is noticeably worse than calibration to

Table 6. Calibration to iTraxx, all maturities simultaneously.

Thresholds (%)	Maturity (yrs)	Market premium (bps)	MinRel premium (bps)	Absolute error (bps)
0–3	5	23.3	21.0	2.3
3–6	5	68.0	71.9	3.9
6–9	5	19.0	25.7	6.7
9–12	5	9.5	14.5	5.0
12–22	5	4.3	7.7	3.4
0–3	7	43.1	43.7	0.6
3–6	7	202.0	208.1	6.1
6–9	7	46.5	53.6	7.1
9–12	7	24.0	27.5	3.5
12–22	7	9.0	12.2	3.2
0–3	10	54.3	56.0	1.7
3–6	10	580.0	590.7	10.3
6–9	10	117.0	127.7	10.7
9–12	10	51.5	54.9	3.4
12–22	10	20.0	22.8	2.8

[14]Data was kindly supplied by Credit Suisse.

single maturity data, which suggests that each maturity has a different underying copula associated with the market tranche quotes. If this is the case then the fitted MinRelEnt copula is in a sense an "average" of the different underlying copulas.

Notwithstanding the poor fit, we also calibrated the MinRelEnt copula to the 5 and 10 year quotes only and priced 7 year tranches out-of-sample to see how well it matched market prices. The results are shown in Table 7. Apart from the super senior tranche — whose predicted premium is double the true value — we can see that the MinRelEnt copula has the ability to interpolate across maturities to within reasonable accuracy, even though overall the calibration results were poor.

In order to investigate further the issue of stationarity of the copula, we generated loss distributions from each of the MinRelEnt copulas calibrated to single maturity data and compared them visually. The *loss distribution* is just the probability distribution of the number of defaults in the portfolio for for a fixed time horizon.[15]

Figure 2 shows the 10 year loss distribution generated from the 10 year Min-RelEnt copula. We include the loss distribution generated from the prior Gaussian copula for comparison. The implied loss distribution in the upper figure is multimodal whereas the loss distribution from the prior Gaussian copula is unimodal. We can see that the market is pricing the risk of a "catastrophic" event corresponding to the mode in the 50–60 defaults region.
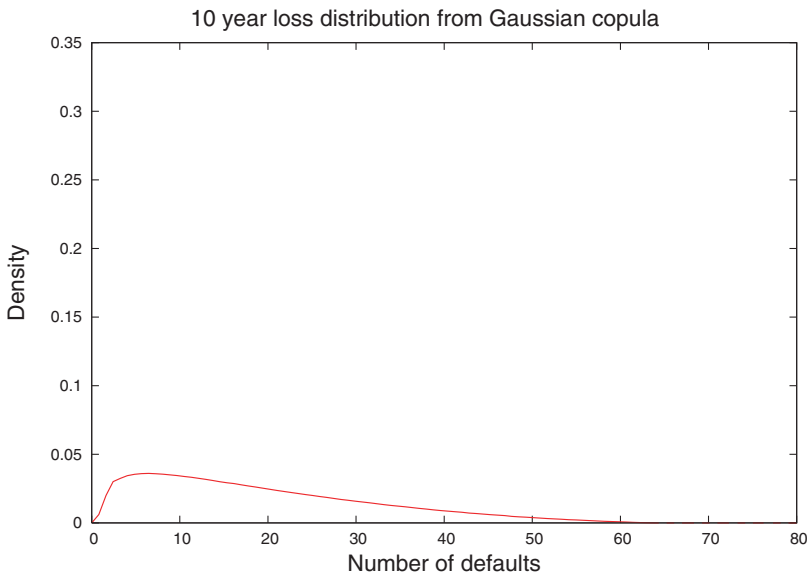
Table 7. Calibration to 5 and 10 year, and pricing 7 year.

Thresholds (%)	Maturity (yrs)	Market premium (bps)	MinRel premium (bps)	Absolute error (bps)
0–3	5	23.3	21.6	2.7
3–6	5	68.0	71.8	3.8
6–9	5	19.0	24.3	5.3
9–12	5	9.5	13.0	3.5
12–22	5	4.3	6.7	2.4
0–3	7	43.1	**42.8**	**0.3**
3–6	7	202.0	**216.8**	**14.8**
6–9	7	46.5	**53.2**	**6.7**
9–12	7	24.0	**25.8**	**1.8**
12–22	7	9.0	**18.1**	**9.1**
0–3	10	54.3	55.3	1.0
3–6	10	580.0	597.9	17.9
6–9	10	117.0	132.5	15.4
9–12	10	51.5	52.3	0.8
12–22	10	20.0	19.2	0.8

---

[15]That is to say it is the plot of the density function for the random variable given by equation (1.1) for a fixed time horizon $t = T$ and $R = 0$.

10 year implied loss distribution
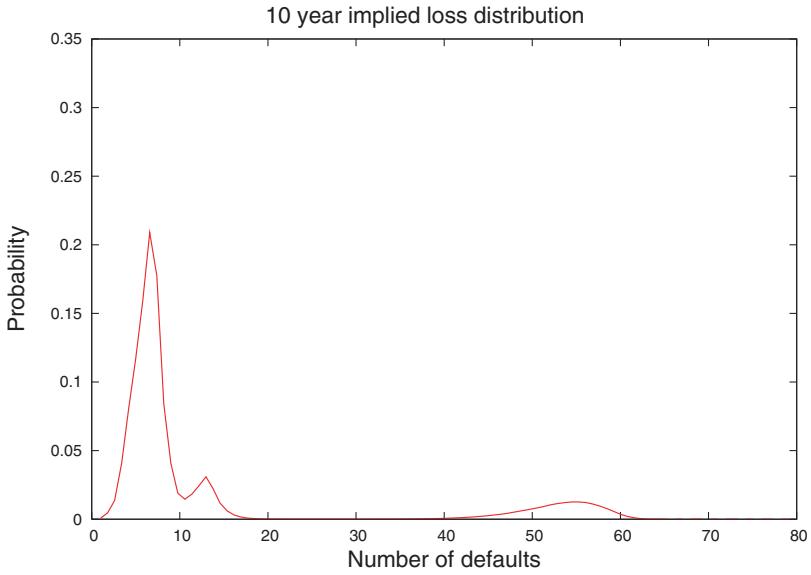
(a) Implied from 10 year MinRelEnt copula

10 year loss distribution from Gaussian copula

(b) Generated from prior Gaussian copula

Fig. 2. Loss distributions for the 10 year horizon.

7 year loss distribution from 10 year MinRelEnt

(a) From 10 year MinRelEnt copula
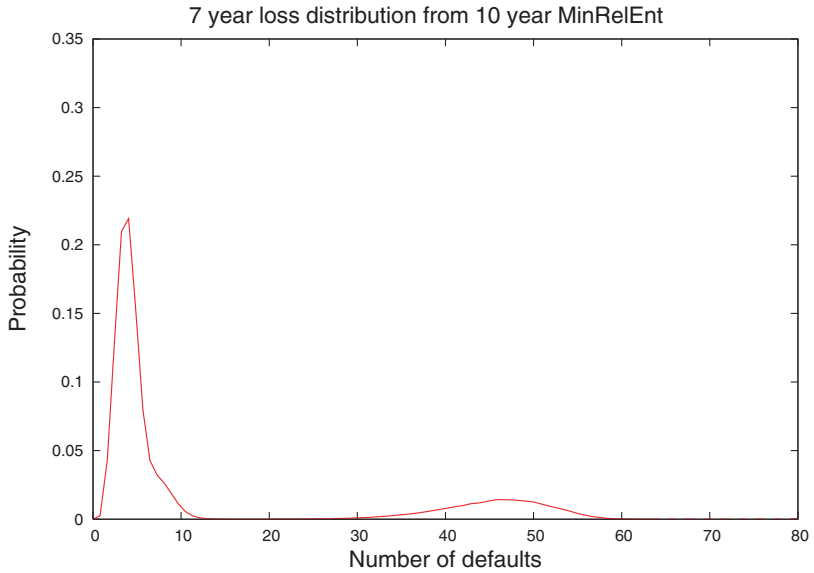
7 year loss distribution from 7 year MinRelEnt
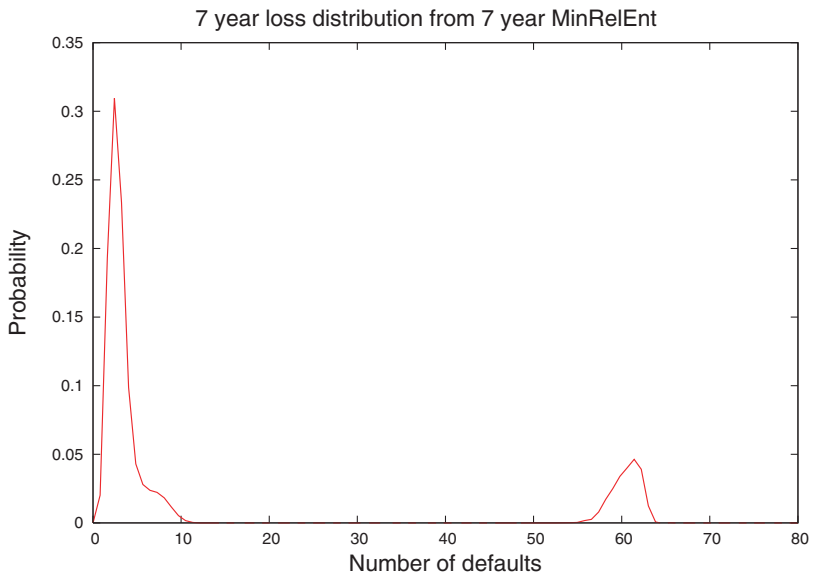
(b) From 7 year MinRelEnt copula

Fig. 3. Implied loss distributions for the 7 year horizon.

5 year loss distribution from 10 year MinRelEnt

(a) From 10 year MinRelEnt copula
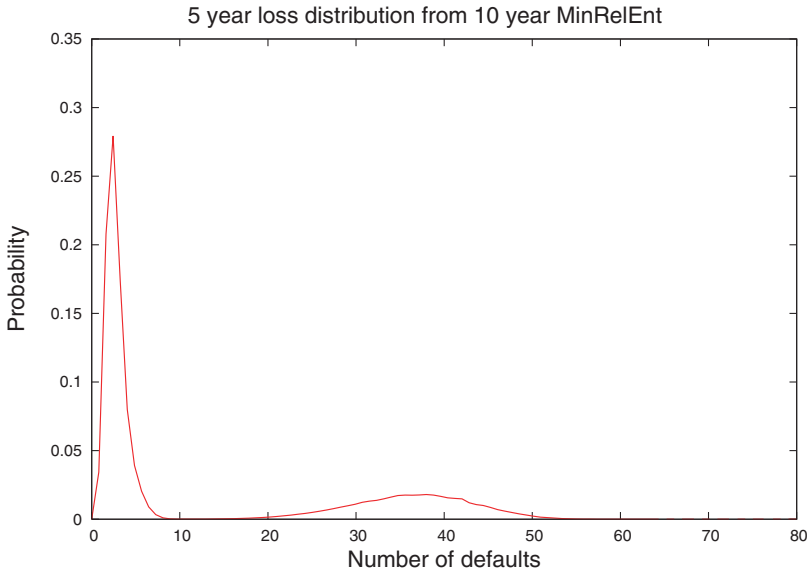
5 year loss distribution from 5 year MinRelEnt
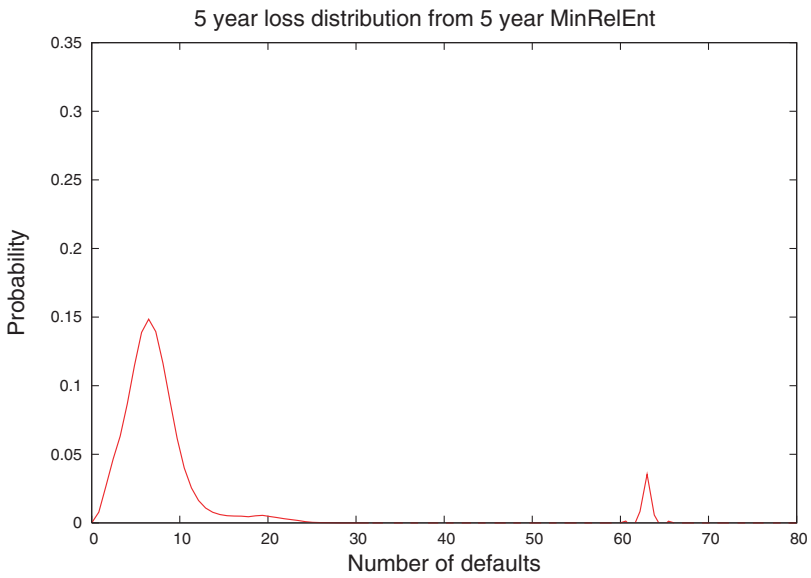
(b) From 5 year MinRelEnt copula

Fig. 4. Implied loss distributions for the 5 year horizon.

We also generate the 5 and 7 year loss distributions from the 10 year MinRelEnt copula, and compare them to the 5 year loss distribution generated from the 5 year MinRelEnt copula and to the 7 year loss distribution generated from the 7 year MinRelEnt copula. This is shown in Figs. 3 and 4.

We can see some difference between the 7 loss distribution implied from the 10 year MinRelEnt copula and the 7 year loss distribution implied from the 7 year MinRelEnt copula, especially in the tail region. The difference is more pronounced in the implied 5 year loss distributions. This suggests that the market view on default dependence changes according to maturity, but in what way over time requires further research.

5. Conclusion

In this paper we have introduced a method to determine the minimum relative entropy copula and applied it to the pricing of CDO tranches. We calibrated the copula to market data, first by using tranches of only one maturity, and then to tranches across different maturities. Although we achieved excellent fits to single maturity data, the fit was noticeably worse for calibration across all maturites. However, when the same exercise was repeated using "market data" simulated from a known copula, we achieved a near-perfect fit across all maturities. Furthermore, we generated loss distributions from the empirical copulas implied from single maturity data and found that, for the same fixed time horizon, they were quite different from each other. These two observations suggest that the market view on default dependence may not be stationary across time.

The advantages of the entropic method are that it provides some justification for the choice of the copula, provides excellent fits to data and performs well out-of-sample. The entropic approach also allows us to empirically investigate whether or not default time dependency remains stationary across time.

There are two main disadvantages however. One is that like most copula methods it is assumed that when used for pricing, the dependence structure between default times remains static over time. As we have seen this may not be the case in reality. The other disadvantage is that both calibration and pricing involves computationally intensive procedures.

The entropic copula method is of course not limited in application to CDO tranche pricing but can be used wherever dependence is involved, e.g. for basket options.

References

[1] Bear Stearns, Valuing and hedging synthetic CDO tranches using base correlations, Technical Report (2004).
[2] J. M. Borwein, R. Choski and P. Maréchal, Probability distributions of assets inferred from option prices via the principle of maximum entropy, *SIAM Journal on Optimization* **14**(2) (2003) 464–478.

[3] C. G. Broyden, The convergence of a class of double rank minimization algorithms, *Journal of the Institute of Mathematics and its Applications* **6** (1970) 76–90 and 222–231.

[4] P. W. Buchen and M. Kelly, The maximum entropy distribution of an asset inferred from option prices, *Journal of Financial and Quantitative Analysis* **31**(1) (1996) 143–159.

[5] X. Burtschell, J. Gregory and J. P. Laurent, A comparative anaylsis of CDO pricing models, Working Paper, BNP Paribas (2005).

[6] S. Das, D. Duffie, N. Kapadia and L. Saita, Common failings: How corporate defaults are correlated, *Journal of Finance* **62**(1) (2007) 55–92.

[7] A. Decarreau, D. Hilhorst, C. Lemaréchal and J. Navaza, Dual methods in entropy maximization: Application to some problems in crystallography, *SIAM Journal on Optimization* **2**(2) (1992) 173–197.

[8] R. Fletcher, A new approach to variable metric algorithms, *Computer Journal* **13** (1970) 317–322.

[9] B. R. Frieden, *Physics from Fisher Information* (Cambridge University Press, Cambridge, 1998).

[10] D. Goldfarb, A family of variable metric methods derived by variational means, *Mathematics of Computation* **24** (1970) 23–26.

[11] G. D. Graziano and L. C. G. Rogers, A dynamic approach to the modelling of correlation credit derivatives using Markov chains, Working Paper, University of Cambridge (2006).

[12] R. J. Hawkins and B. R. Frieden, Financial probabilities from Fisher information, eprint, arXiv:cond-mat/0302579 (2006).

[13] J. Hull and A. White, Valuing credit derivatives using an implied copula approach, *Journal of Derivatives* **14**(2) (2006) 8–28.

[14] J. C. Jackwerth and M. Rubinstein, Recovering probability distributions from option prices, *Journal of Finance* **51** (1996) 1611–31.

[15] E. T. Jaynes, Information theory and statistical mechanics, *Physical Review* **106** (1957) 620–630.

[16] M. S. Joshi, Applying importance sampling to pricing single tranches of CDOs in a one-factor Li model, *Wilmott* **March** (2005).

[17] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth and A. H. Teller, Equations of state calculations by fast computing machines, *Journal of Chemical Physics* **21**(1) (1953) 1087–1092.

[18] A. Mortensen, Semi-analytical valuation of basket credit derivatives in intensity-based models, *Journal of Derivatives* **13**(4) (2006) 8–26.

[19] D. F. Shanno, Conditioning of quasi-Newton methods for function minimization, *Mathematics of Computation* **24** (1970) 647–657.