

MINIMAL MODELS FOR CREDIT RISK : AN INFORMATION THEORY APPROCH

Vivien BRUNEL

Risk Department
Société Générale
92972 Paris La Défense cedex, France
tel : +33 1 42 14 87 95
email : vivien.brunel@socgen.com

09/07/04

Abstract

Statistical estimation is one of the classical methods for assigning numerical values to probabilities, especially in the field of financial engineering. However, in the case of credit risk applications, empirical data are often of poor quality, and it is difficult to choose the model assumptions. The principle of maximum entropy comes from information theory and aims at inferring probability distributions under the constraint of the available information. This method, when applied to credit risk, leads to the models containing the minimal assumptions coherent with the available information. These models are called minimal models and constitute a reference point when we design a new model. We apply this approach to the choice of the loss distribution of credit portfolios, asset backed securities and to the distribution of recovery rates, and we show how to use it for challenging the assumptions of any model in these areas.

Acknowledgements : this article has benefited from insightful comments from Richard Dalaud and from the two anonymous referees. I also acknowledge Mikael Bouvier, Mathieu L'hoir and Morgane Prou for their helpful reading of the manuscript.

1. INTRODUCTION

It is commonly admitted that part of the difficulty in modeling credit risk comes from the fact that credit markets are not yet as organized and liquid as stock, currencies and interest rates markets. As a result, there is no reference model for credit risk contrary to the other types of markets that benefit the Black-Scholes seminal breakthrough. Moreover, data for credit risk are very difficult to deal with, simply because credit events are rare. However, quantitative approaches for credit risk are a major subject of interest in the financial community. There are two reasons for this. First, credit markets are getting more and more liquid and provide new ways of funding and investing. Second, the Basle Committee gives to the financial institutions a strong incentive for a better knowledge of their credit portfolios and for a more accurate management of their credit exposures. This is why quantitative methods for credit derivatives, spread dynamics and portfolio modeling are at the core of many recent developments in the financial literature.

The lack of reliable empirical data is a challenge to justify and calibrate a credit model. Many credit risk models are based on assumptions that are not always well justified from empirical and statistical perspectives. The information we have about the drivers of the occurrence of credit events is generally

partial and incomplete. For instance, when we consider a loan, we know the name and the rating of the counterparty, but generally, very few additional data are available.

Instead of focusing on statistical approaches or on inaccurate models for credit risk, and since no reference model similar to Black-Scholes' model exist, we propose here to use an approach based on information theory. The principle of maximum entropy (MAXENT) is a method of statistical inference which aims at assigning a numerical value to the probability of occurrence of a random event under the constraint of the available information. This method comes from physics and dates back to the 19th century. Boltzmann and Gibbs discovered its probabilistic nature and entropy has become a cornerstone concept in statistical physics as shown in [8]. In 1948, Shannon [12] exhibited the relevance of the entropy measure in the field of communication theory and he launched the foundations of a new science that has led to hundreds of applications in many other areas. In the field of finance, for instance, the entropy method has been used for model calibration [1], incomplete markets [3], and more recently, in the field of model performance measure [2]. We find two other examples in the context of credit portfolio modeling, Thomson and Ordovas [13] calculate the average contribution of each line of a portfolio to the total loss. They approximate this as the most likely default rate of each line conditionally to the value of the portfolio loss. They are then led to maximize an entropy under the constraint of a fixed total portfolio loss. The second example deals with the loss distribution of a homogeneous portfolio when both the expected loss and default correlation are known. Molins and Vives [10] solve the problem thanks to the MAXENT technique and compute the tail distribution of a credit portfolio.

In this paper, we use the entropy method as a statistical inference tool applied to several topics : loss distribution of credit portfolios, random recovery rates, loss distribution of structured transactions. For our purpose, the MAXENT method is not a meta-model that would assign numerical values where the other methods have failed ; the goal of the MAXENT method is to provide a basis for challenging the assumptions of the model we retain.

In section 2, we introduce Jaynes' formalism for statistical inference. This method is applied to structural credit models in section 3. In section 4, we apply this method to the recovery rate distribution function and we compare it to the use of the beta distribution. In section 5, we consider the problem of the loss distribution for Asset-Backed Securities (ABS). Finally, in section 6, we propose other possible applications in the field of credit risk management, and we conclude.

2. JAYNES' FORMALISM

Let us consider a random experiment with N possible outcomes, and let p_i be the probability of outcome i . We also call $p = (p_1, \dots, p_N)$ the whole probability distribution of the random experiment. If the probability distribution is peaked on a single outcome j (*i.e.* $p_j = 1$ and $p_i = 0$ for all $i \neq j$), then there is no uncertainty about the outcomes of the experiment since we are sure with probability 1 of the occurrence of outcome j . On the other hand, if p is the uniform distribution over the N possible outcomes (*i.e.* $p_i = 1/N$ for all i), all the outcomes are equi-probable, and the uncertainty about the result of the experiment is maximal. In both example, we say that the informational content of the probability distribution reaches a maximum or a minimum respectively. For any other probability distribution, the informational content is in between the two previous cases. In his seminal work about information theory, Shannon [12] has found a functional measure of the informational content of a probability distribution. By imposing a set of a few natural requirements on this functional measure (continuity in the p_i , maximum obtained for the uniform distribution, and independence relative to the gathering of random events) he showed that the only relevant functional was the entropy functional :

$$S(p) = - \sum_{i=1}^N p_i \ln p_i \quad (1)$$

In the continuous limit, we can easily extend the principle of maximum entropy for a probability distribution $p(x)$ by considering the functional :

$$S(p) = -\int p(x) \ln p(x) dx \quad (2)$$

Jaynes [5,7] has used the principle of maximum entropy as a method for assigning numerical values to probabilities when certain partial or incomplete information is available. More precisely, suppose that the probabilities $p = (p_i)_{1 \leq i \leq N}$ are assigned to a set of N mutually exclusive outcomes of an experiment. Jaynes' principle states that these probability values are to be chosen so that the Shannon entropy $S(p)$ reaches a maximum, under the condition that the probability distribution p agrees with the available information. The intuition behind Jaynes' approach is clear : in the process of assigning numerical values to the probabilities of a random experiment, we look for the less informative distribution (MAXENT distribution) under the constraint of the available information about the outcomes. We illustrate the principle by a famous example taken from [6]. Suppose that a dice is tossed a large number of times and we are told that the average number of spots up is not 3.5 as we might expect from a "true" die, but 4.5. Mathematically, we interpret this through the constraint (with $N = 6$) :

$$\sum_{i=1}^N i p_i = 4.5 \quad (3)$$

Given this information and nothing else, what estimate could we make about the probabilities p_i with which the different faces of the dice appear ? We solve the MAXENT problem with two constraints :

$$\left\{ \begin{array}{l} \max_{(p_i)_{1 \leq i \leq N}} \left(-\sum_i p_i \ln p_i \right) \\ u.c. \left\{ \begin{array}{l} \sum_i i p_i = 4.5 \\ \sum_i p_i = 1 \end{array} \right. \end{array} \right. \quad (4)$$

To this end, we introduce two Lagrange multipliers λ_1 and λ_2 to take the constraint into account. The Lagrange function then writes :

$$L(p, \lambda_1, \lambda_2) = S(p) - \lambda_1 \left(\sum_i i p_i - 4.5 \right) - \lambda_2 \left(\sum_i p_i - 1 \right) \quad (5)$$

Kuhn-Tucker's theorem states for the existence and uniqueness of the solution. The optimality conditions are given by the relations :

$$\frac{\partial L}{\partial p_i} = 0 = \frac{\partial L}{\partial \lambda_1} = \frac{\partial L}{\partial \lambda_2} \quad (6)$$

This leads to the MAXENT probability distribution for this problem :

$$p^* = (5.43\%, 7.88\%, 11.42\%, 16.54\%, 23.98\%, 34.75\%) \quad (7)$$

This distribution has entropy $S^* = 1.6138$ compared to the value $\ln 6 = 1.7918$, corresponding to the case of no constraint and uniform distribution. The distribution with 50% of outcomes 3 and 50% of outcomes 4 also satisfies the mean value constraint but has an entropy equal to $\ln 2 = 0.6931$.

The entropy criterion provides us with a way to select a probability distribution under constraint. We may now question to what extent the MAXENT distribution is preferred to other distributions with lower entropy. In particular, when we make n random experiments, what is the proportion of empirical distributions that are satisfying the constraints in a range of entropy of size ΔS around the maximum

entropy? Jaynes' concentration theorem states that nearly all the sequences satisfying the constraints have empirical frequencies extremely close to the MAXENT probabilities. For instance, in the above die example, if we make 10000 tosses of the die, sequences with 50% of outcomes equal to 4 and 50% of the outcomes equal to 5, satisfy the mean value constraint, but are extremely rare and are leading to an entropy far from the maximum entropy. Jaynes [6] answers the question in the large n limit, and shows that $2n\Delta S$ is distributed as a chi-squared distribution with $k = N - m$ degrees of freedom, where N is the number of possible occurrences in the random experiment and m is the number of constraints entering the optimization program (cf. eq.(4), $N = 6$ and $m = 2$ in this case). In the upper die example taken from Jaynes [6], let us consider $n = 1000$ experiments. Applying the concentration theorem, we have $k = 6 - 2$ degrees of freedom and we find that 95% of all the possible empirical distributions satisfying the constraints in eq. (4) have entropy in the range $\Delta S = (2n)^{-1}(\chi_4^2)^{-1}(0.05) = 0.00474$, where χ_4^2 is the chi-squared distribution function with 4 degrees of freedom. As the maximum value of the entropy is equal to $S^* = 1.6138$, we state that 95% of the outcomes have entropy in the range $1.609 \leq S \leq 1.6138$. If we are given another distribution function satisfying the constraints, the computation of its entropy provides us with a criterion for rejecting it or not relative to the MAXENT distribution up to a given confidence level. For example, this criterion would lead to a rejection of the bimodal distribution $p = (0,0,50\%,50\%,0,0)$.

The main features of the concentration theorem are :

- The size of the range doesn't depend on the value of the maximum entropy, only on the number of degrees of freedom.
- The size of the range decreases asymptotically as $1/n$ and not as $1/\sqrt{n}$ as it is generally the case in probability asymptotic results. Thus, the entropy range of "admissible" distributions decreases with n faster than expected.

We now have a tool for assigning probabilities by taking into account only the available information. We may go even further ; if we make assumptions about the modeling of a system, we are then able to build the minimal model coherent with these assumptions. Moreover, if a set of models is given, we are able to choose the model associated with the less restrictive set of assumptions, *i.e.* the model that has the less informative content. This is what we aim at doing in the next section in the field of credit risk for homogeneous credit portfolios.

3. THE MINIMAL STRUCTURAL MODEL

The structural approach for credit risk is based on the financial structure of the firms. In the most simple approach, first proposed by Merton [9], a firm defaults on its debt at maturity if the asset value is not sufficient to pay back the debt. In this model, the asset returns are assumed to have normal distributions, and when we consider a pool of firms, the underlying assumption is that the asset returns have multivariate normal distributions. This model is commonly used in the credit risk industry (CreditMetrics for instance) because it is easy to implement and to calibrate, it leads to a deep understanding of portfolio effects. Moreover, from a mathematical viewpoint, this model provides a closed form formula for the loss distribution in the case of homogeneous fine-grained portfolios ([14]).

The starting point of our analysis is the MAXENT problem of a multivariate distribution with constraints on the mean vector and on the covariance matrix. Let X be a random vector with mean vector m and covariance matrix Σ . The MAXENT distribution f for the random vector X is constrained by the three following relations :

$$\left\{ \begin{array}{ll} \int f(X) dX = 1 & \text{(Normalisation constraint)} \\ \int X f(X) dX = m & \text{(Expected value constraint)} \\ \int X^T X f(X) dX = \Sigma & \text{(Variance - covariance constraint)} \end{array} \right. \quad (8)$$

The solution of the MAXENT problem with this system of constraints, leads to the normal multivariate distribution :

$$f(X) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left[-\frac{(X-m)^T \Sigma^{-1}(X-m)}{2}\right] \quad (9)$$

We consider a homogeneous pool of N loans, issued by N firms that have the same default probability p . The homogeneity assumption also includes that the assets of the firms have the same pair-wise correlation $r > 0$. To make the assumptions more precise, the existence of pair-wise correlations requires that the asset returns have a finite standard deviation. For simplicity, we thus assume that the asset returns all have an expected value equal to 0, a standard deviation equal to 1 and pair-wise correlations equal to r . As we have just seen, the MAXENT distribution for the asset returns in this case is the normal multivariate distribution. This provides Vasicek's approach [14] of credit portfolios with an interesting informational interpretation since we have shown that Merton's multivariate model is the minimal structural model. In the limit $N \rightarrow \infty$ (infinitely granular portfolio), the minimal structural model leads to Vasicek's cumulated distribution for the losses of a homogeneous portfolio :

$$P(L < x) = N\left[\frac{\sqrt{1-r}N^{-1}(x) - N^{-1}(p)}{\sqrt{r}}\right] \quad (10)$$

where the function $N(\cdot)$ is the cumulative normal distribution. We note that contrary to [10], we assume here a known asset correlation instead of default correlation. This assumption leads to different results in the MAXENT paradigm.

The next step in our study is to relax the assumption of known uniform asset correlations, and to assume that the asset correlations are all equal but that their value is unknown. The choice of the value of r is not neutral because it corresponds to different values of the entropy, and then leads to loss distributions that don't have the same informational content. The puzzle is solved by choosing the MAXENT distribution of the form described in eq.(10) when r varies. In Figure 1, we show how the entropy of the Vasicek distribution varies with the asset correlation for a default probability $p = 0.05\%$. We see that the entropy equals to 0 in the two extreme cases where the correlation is equal to 0 and 100 % respectively, because the portfolio loss distribution is peaked. The maximum entropy is reached for an asset correlation equal to $r^* = 7.78\%$. The meaning of this figure appears when we compare it to other correlation assumptions : for instance, assuming that the asset correlation for a given asset class (with an average default rate of 0.05%) is around 20%, is a quite strong assumption, since the minimal hypothesis (given by the MAXENT formalism) would be to assume a 7.78% correlation level. Then, comparing the assumption we make to the minimal entropic results leads to a quantitative basis of challenging the strength of our assumption ; in the above example, the 20% correlation assumption needs to be justified by some additional arguments, for instance, economical, historical or financial arguments.

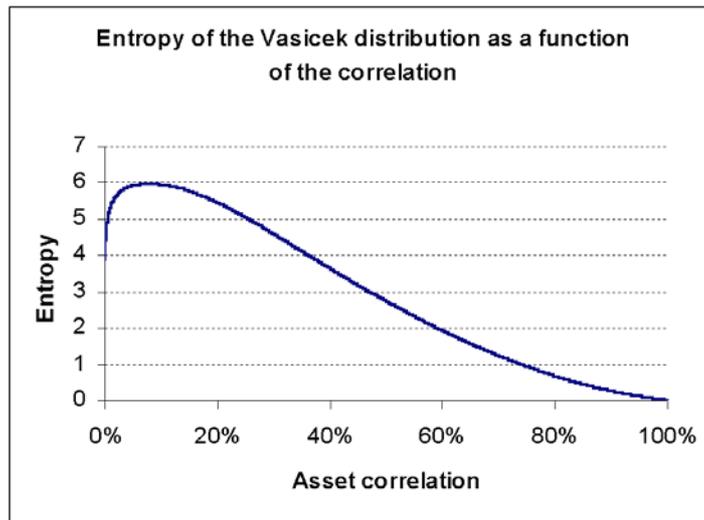


Figure 1 : Entropy of the Vasicek distribution ($p = 0.05\%$) as a function of correlation

Figure 2 gives the value of the MAXENT correlations for the Vasicek distribution as a function of the default probability p . This curve is increasing in the area of realistic obligor default probabilities (*i.e.* between 0% (AAA rating) and around 20% (CCC rating)) conversely to the regulatory curves proposed for large corporates and “other retail exposures” by the Basel Committee in June 2004. Here again, we have the illustration that the Basel Committee may have strong economical (and maybe political) incentives to assume a decreasing (or flat for mortgages and revolving retail exposures) relationship between correlation and default probability whereas the minimal (MAXENT) behavior would be an increasing relationship.

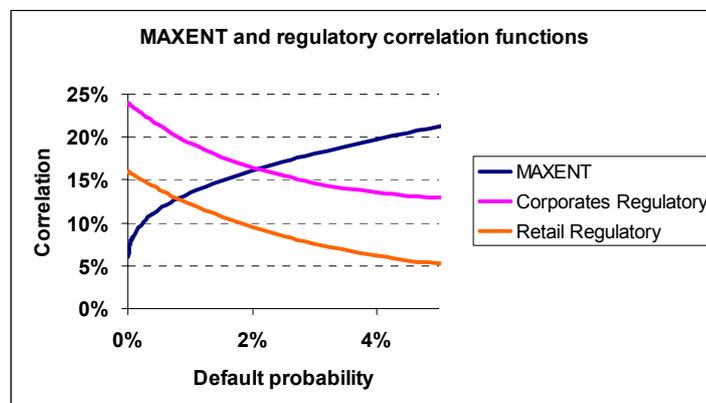


Figure 2 : MAXENT and regulatory (large corporates and “Other retail exposures” as in the June 2004 Basle document) asset correlation functions, as a function of the default probability.

4. RANDOM RECOVERY RATE

In this section, we deal with the stand alone distribution of recovery rates, without any correlation with the default event. In the case of recovery rates, we often have an estimate of the average recovery rate, but the question of the probability distribution of the recovery rate is much more difficult and still controversial. Many approaches have been proposed, but, in practice, we often use a beta law to model random recoveries, in spite of a lack of conceptual or empirical evidence to do so. Recently, there have been some more “structural” models to arrive at the recovery rate distribution starting from an

assumption on the distribution of the collateral value (normal distribution in Frye [4] and lognormal in Pykhtin [11]).

The principle of maximum entropy provides with another answer when the expected recovery rate μ is given. The recovery rate is a random variable taking its value in the interval [0%, 100%]. It leads to the following MAXENT problem :

$$\begin{aligned} \max_{p(x)} & \left[- \int_0^1 p(x) \ln p(x) dx \right] \\ \text{u.c.} & \left\{ \begin{array}{l} \int_0^1 x p(x) dx = \mu \quad (\text{Expected value constraint}) \\ \int_0^1 p(x) dx = 1 \quad (\text{Normalisation constraint}) \end{array} \right. \end{aligned} \quad (11)$$

In this optimization program, we implicitly assume that the probability of having exactly 0% or 100% recoveries are equal to 0 ; we shall go beyond this assumption a bit below, in the next section. After writing an optimum equation similar to eq.(6), we show that this optimization problem has a closed form solution : the truncated exponential distribution :

$$p(x) = C e^{-\lambda x}, \quad x \in [0,1] \quad (12)$$

where C and λ are chosen to satisfy the constraints in eq. (11) :

$$C = \frac{\lambda}{1 - e^{-\lambda}} \quad \text{and} \quad \mu = \frac{1}{\lambda} - \frac{e^{-\lambda}}{1 - e^{-\lambda}} \quad (13)$$

Eq. (13) can be solved numerically. It is interesting to compare this truncated exponential law with the beta law that is generally used to model recovery distributions. The very reason why the beta law is extensively used in the modeling of recovery laws is because the variety of shapes induced by the beta function is very wide. Moreover, as described in eq. (14), the beta function is a two parameters law and is easily expressed in terms of expected value and standard deviation ; this makes the calibration relatively straightforward. The beta law is the law of a random variable X that belongs to the interval [0,1] and with a density function :

$$f(x) = \frac{1}{B(n, p)} x^{n-1} (1-x)^{p-1} \quad \text{with } n, p > 0 \quad \text{and} \quad B(n, p) = \frac{\Gamma(n)\Gamma(p)}{\Gamma(n+p)} \quad (14)$$

We can express the expected value and the variance of X in terms of the parameters n and p :

$$E[X] = \frac{n}{n+p} \quad V[X] = \frac{np}{(n+p+1)(n+p)^2} \quad (15)$$

We have represented in figure 3 the MAXENT distribution function for an expected recovery rate equal to 70% and the beta function calibrated on this expected recovery rate and on the standard deviation of the recovery rate of the MAXENT distribution. The parameters for the MAXENT distribution are $C = 0.20$ and $\lambda = -2.67$, and for the beta law, $n = 1.68$ and $p = 0.72$. Both distributions have expected value 70% and standard deviation 24.85%.

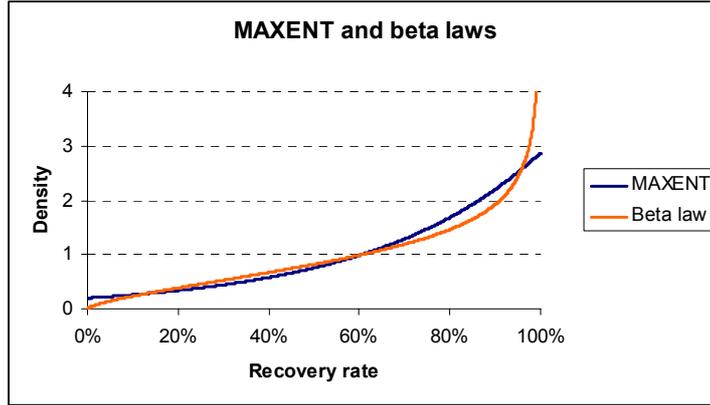


Figure 3 : MAXENT law for an expected recovery rate of 70% and beta law with the same first two moments

We see graphically that these distributions are close to each other, but Jaynes' concentration theorem discussed in section 2 is going to provide us with a more quantitative judgement. In the continuous limit, there is no extension of the concentration theorem but we don't need it since when the number of random experiments n is finite, the empirical distribution function is discrete. We are then going to apply the concentration theorem after making a discretisation of the MAXENT distribution into an histogram of N blocks, with of course the condition that N is much smaller than the number of random experiments n . For numerical purposes, we are going to consider $n = 10000$ and $N = 100$. The concentration theorem writes :

$$2n\Delta S = \left(\chi_{N-2}^2\right)^{-1}(1-q) \quad (16)$$

We show that a fraction q of the empirical distributions satisfying the expected value constraint have an entropy in the interval $\Delta S = \left(\chi_{98}^2\right)^{-1}(1-q)/2n$. For $q = 99\%$, we get $\Delta S = 0.0067$, meaning that 99% of the distribution functions satisfying the constraints have an entropy in the range of width 0.0067 below the maximum entropy value. When we compute the entropy of the beta law, we see that it has an entropy equal to $S_{\max} - 0.025$, far outside of the 99% confidence interval provided by the concentration theorem.

As the beta law is a two parameters law, it is legitimate to compare the MAXENT results to the beta function when both the expected value and the standard deviation of the recovery rate are given. In this case, the maximum entropy distribution is the truncated normal distribution on the interval $[0,1]$. The optimization program is the one-dimensional version of eq. (8) and then leads to a normal distribution as in eq. (9). Since the recovery is constrained to stay in the range $[0,1]$, this normal distribution is truncated, and the parameters are computed numerically from the expected value and the standard deviation. Figure 4 shows the MAXENT distribution and the beta distribution with the same mean value and standard deviation equal to 70% and 18% respectively.

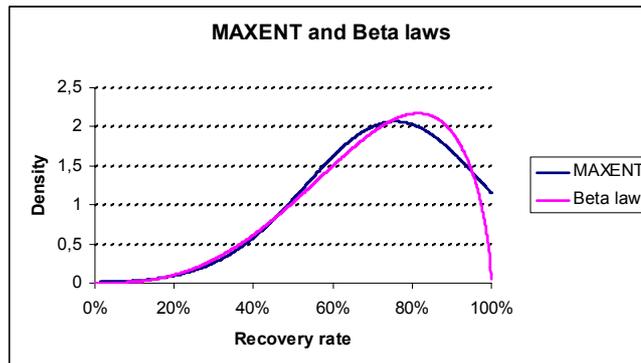


Figure 4 : MAXENT and beta laws with the same mean value and standard deviation
(equal to 70% and 18% respectively)

As above, we apply the concentration theorem, but there is an additional constraint compared to the previous case, and the number of degrees of freedom entering eq.(16) is then 97. This leads to a slight change in the width of the 99% confidence interval : $\Delta S = 0.0066$. The computation of the entropy of the beta distribution compared to the entropy of the truncated normal distribution (the maximum entropy) leads to an entropy of the beta law equal to $S_{\max} - 0.013$, which is again far outside of the 99% confidence interval.

In the last two examples, we have shown that when we run 10000 random experiments and try to infer a distribution function over 100 blocks, the choice of the beta law relative to the MAXENT distribution is not relevant up to a 99% confidence level (and even less). This is a strong incentive for using the MAXENT formalism.

At the beginning of this section, we have assumed that the probability of having 0% or 100% recovery was equal to 0. This is often a reasonable assumption when we consider the recovery on a defaulted corporate firm for a senior unsecured traditional banking transaction. On the reverse this assumption is no longer correct in a general framework of recovery, especially when a collateral enters the transaction as a mitigant of credit risk or when we consider credit exposures on asset-backed securities. In the case of recoveries contingent to the collateral value at default, the transaction may be over-collateralised or on the reverse, the collateral value may be equal to zero. As a consequence, we need to add two mass points in the recovery rate distribution located at 0% and 100%. The case of the ABS distribution losses is strictly similar to this situation, and is treated in the next section.

5. LOSS DISTRIBUTION OF ABS TRANCHES

The question of inferring the loss distribution of an ABS is very similar since we are looking for a distribution function with support $[0,1]$, given some information (most of the information available comes from the rating). The rating of the ABS leads to a constraint about the expected loss thanks to the rating agencies' (for instance Moody's or Fitch) correspondence tables between rating and expected loss. However, the constraints on the distribution function are a bit different compared to the previous example, according to the seniority of the tranche we are considering. For instance, the underlying pool may have a finite number of exposures leading to two mass points in the loss distribution for $x=0$ and $x=1$, meaning that there is a non zero probability to lose 0% or 100% of the exposure. An other possibility, is that we are dealing with a mezzanine note, and the probabilities that the losses on the ABS are 0 or 100% are both non zero. In this case, the loss distribution writes :

$$f(x) = p_0 \delta(x) + p(x) + p_1 \delta(x-1) \quad (17)$$

The MAXENT approach is very relevant for a bank investing in ABS because the underlying pool of the ABS is generally not very well known : the ABS is often rated by a rating agency that gives a reliable estimate of its credit quality, but very low information about the underlying pool and the size of the credit enhancement is available. In many cases, Vasicek's model (see section 3) is difficult to calibrate because of the lack of information about the underlying assets, and also because the assets are not homogeneous and the pool not granular ; Collateralised Debt Obligations (CDO), Collateralised Loan Obligations (CLO) and Collateralised Bond Obligations (CBO) illustrate perfectly the need of the MAXENT approach. We point out that even if the pool is not very granular, it is relevant to seek a continuous MAXENT loss distribution because the LGD of the loans are unknown and/or random. The optimization problem writes :

$$\max_{p_0, p_1, p(x)} \left[\begin{aligned} & -p_0 \ln p_0 - p_1 \ln p_1 - \int_0^1 p(x) \ln p(x) dx - \\ & -\lambda_1 \left(p_1 + \int_0^1 x p(x) dx - \mu \right) - \lambda_2 \left(p_0 + p_1 + \int_0^1 p(x) dx - 1 \right) \end{aligned} \right] \quad (18)$$

where p_0 (resp. p_1) is the probability of zero loss (resp. 100% loss) on the ABS tranche, and μ is the expected loss of the ABS tranche. As in eq. (5), we solve the optimization program by differentiating the lagrangian function of eq.(13) with respect to p_0 , p_1 , $p(x)$ and the Lagrange multipliers λ_1 and λ_2 . The solution of this optimization problem is again a truncated exponential law with two additional mass points at $x=0$ and $x=1$.

$$p(x) = p_0 e^{-\lambda x}, \quad p_0 = \lim_{x \rightarrow 0} p(x), \quad p_1 = \lim_{x \rightarrow 1} p(x) \quad (19)$$

where the parameters are the solution of the constraint equations :

$$\begin{aligned} p_0 \left(1 + e^{-\lambda} + \frac{1 - e^{-\lambda}}{\lambda} \right) &= 1 && \text{(Normalisation constraint)} \\ p_0 \left(e^{-\lambda} + \frac{1 - e^{-\lambda}}{\lambda^2} - \frac{e^{-\lambda}}{\lambda} \right) &= \mu && \text{(Expected loss constraint)} \end{aligned} \quad (20)$$

We consider the example of a mezzanine ABS tranche rated BB-, corresponding to a yearly expected loss of 1.20%. In figure 5, we have the MAXENT loss distribution of such a tranche. In particular, the probability of no default over one year is equal to 89.7%, and the probability of a 100% loss is close to 0.

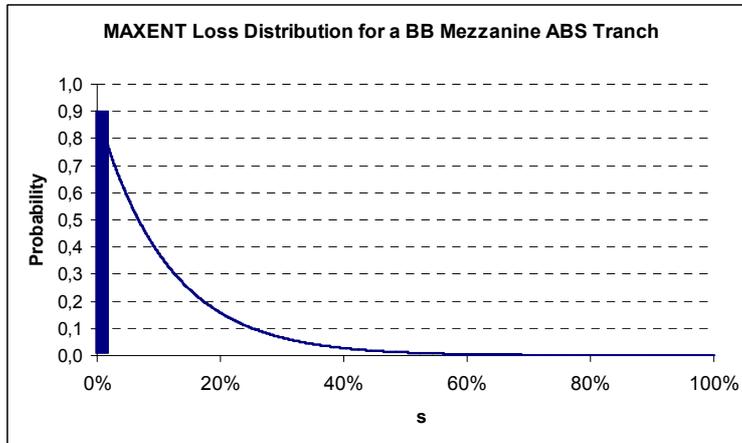


Figure 5 : MAXENT loss distribution for a mezzanine ABS tranche with a BB rating

The above analysis done for ABS, could be applied straightforwardly to the case of random recovery rate because the mathematical framework is exactly the same. In the case of ABS however, the expected loss constraint as described above is not completely realistic since the rating does not correspond exactly to an expected loss but to an interval of expected losses. For instance a rating BB corresponds to a one-year expected loss in the range $[0.86\%, 1.55\%]$. It is possible to insert this constraint in eq.(18) by introducing two Lagrange multipliers instead of one for the expected loss constraint as this constraint can be expressed with two inequalities instead of one equality. Kuhn-Tucker's theorem still applies and the optimization process selects the expected loss value in the constraint interval so as to maximize the entropy. Let's call $[\mu_0, \mu_1]$ the range of admissible expected losses. If $\mu_1 < 50\%$ then μ_1 is selected (this is the most common case). If $\mu_0 > 50\%$ then μ_0 is selected. If $\mu_0 < 50\% < \mu_1$, then the value 50% is selected. This result is quite intuitive because we know that the entropy is maximum in the case of an

expected loss equal to 50% (uniform probability distribution) ; this is why the selected value of the expected loss in the constraint interval is the closest value to 50% that belongs to the admissible interval.

Another possibility is that there is a constraint on the value of p_0 which is the probability of default of the ABS tranche. For instance we know that S&P ratings correspond to default probabilities rather than expected losses. This constraint can easily be taken into account in the optimization program in eq.(18), leading to another truncated exponential formulation for the solution of the problem.

6. CONCLUSION

The principle of maximum entropy is an interesting technique for assigning numerical values to probabilities for two main reasons. The first reason is that it does not take into account any assumption additional to the available information. This leads to a minimal model from the point of view of the information content. The second reason is that it provides us with a reference model, useful for choosing some key parameters such as the asset correlation for credit portfolios for instance, or quantifying the weight of additional information in a given model.

We applied successfully this criterion to credit risk. First, we have shown that in the structural approach of credit portfolio, the value of the asset correlation at fixed default probability is an important driver of the informational content of the loss distribution. It is possible to choose the parameters so as to maximize the entropy in the structural approach, and we have made a fruitful comparison with the numerical values retained by the Basel Committee. In the second application of the principle of maximum entropy, we have inferred the recovery rate distribution, and compared it to the beta distribution. Jaynes' concentration theorem has led us to the conclusion that the MAXENT distribution was more relevant to use than the beta distribution. Finally, we have applied the entropy formalism to the loss distribution of structured products such as ABS tranches. In particular, it has proved to be very useful in the case when the distribution function has accumulation points. In all these cases of unreliable or insufficient information, the maximum entropy method leads to the minimal hypothesis. The minimal set of hypothesis is a useful reference from which any assumption has to be discussed. Moreover the MAXENT method is useful to classify models from an information content point of view, and the minimal model serves as a reference model when little information is available.

The principle of maximum entropy can be very useful in many other problems. It often happens that a random variable enters our model but we only have little information about it (for instance, we know its expected value and its range). Section 4 illustrates this, in the case of the probability law of the random recovery rate. Many fields of finance are likely to use this entropy method, especially for the quantitative modeling of a business line or any activity that lacks empirical data. An other important potential field of application of this methodology, is risk integration. The most famous problem of risk integration is to infer a dependence model between credit, market and operational risks. We generally know the marginal risks, but the way we should integrate them together is still very challenging, and is a major concern in the field of risk management and banking supervision.

REFERENCES

- [1] Avellaneda M. (1998) : Minimum Relative Entropy Calibration of Asset Pricing Models, International Journal of Theoretical and Applied Finance, Vol 1, N°4, 447-472.
- [2] Friedman C. and Sandow S. (2003) : Model performance measures for expected utility maximizing investors, International Journal of Theoretical and Applied Finance, Vol 6, N°4, 355-401.

- [3] Frittelli M.(2000) : The minimal entropy martingale measure and the valuation problem in incomplete markets, *Mathematical finance*, 10 (1), 39-52.
- [4] Frye J. (2000) : Collateral damage, *Risk*, April, 91-94.
- [5] Jaynes E.T. (1957) : Information theory and statistical mechanics, *Phys. Rev.* 106, 620 and 108, 171.
- [6] Jaynes E.T. (1979) : Concentration of distributions at entropy maxima, 19th NBER-NSF seminar on Bayesian statistics, Montreal.
- [7] Jaynes E.T.(2003) : The logic of science, Cambridge.University Press.
- [8] Landau L. and Lifschitz E. (1980) : Statistical physics, part 1, Butterworth-Heinemann, 3rd edition.
- [9] Merton R. (1974) : On the pricing of corporate debt : the risk structure of interest rates, *Journal of finance*, 29, 449-470.
- [10] Molins J. and Vives E. (2004) : Long range Ising model for credit risk modeling in homogeneous portfolios, <http://arxiv.org/abs/cond-mat/0401378>.
- [11] Pykhtin M. (2003) : Unexpected recovery risk, *Risk*, August, 74-78.
- [12] Shannon C. (1948) : A mathematical theory of communication, *The Bell system technical journal*, Vol. 27, pp.379-423, 623-656, July, October.
- [13] Thomson K. and Ordovas R. (2003 a and b) : Credit ensembles, *Risk*, April, 67-72 ; The road to partitions, *Risk*, May, 93-97.
- [14] Vasicek O. (1991) : Loan loss probability distribution, KMV Corporation.