Consistent Information Multivariate Density Optimizing Methodology

Miguel Angel Segoviano Basurto*
Financial Markets Group
msegoviano@imf.org

March 17, 2006

Abstract

The estimation of the profit and loss distribution of a loan portfolio requires the modelling of the portfolio's multivariate distribution. This describes the joint likelihood of changes in the credit-risk quality of the loans that make up the portfolio. A significant problem for portfolio credit risk measurement is the greatly restricted data that are available for its modelling. Under these circumstances, convenient parametric assumptions are frequently made in order to represent the nonexistent information. Such assumptions, however, usually do not appropriately describe the behaviour of the assets that are the subject of our interest, loans granted to small and medium enterprises (SMEs), unlisted and arm's-length firms. This paper proposes the Consistent Information Multivariate Density Optimizing Methodology (CIMDO), based on the cross-entropy approach, as an alternative to generate probability multivariate densities from partial information and without making parametric assumptions. Using the probability integral transformation criterion, we show that the distributions recovered by CIMDO outperform distributions that are used for the measurement of portfolio credit risk of loans granted to SMEs, unlisted and arm's-length firms.

Key Words: Portfolio Credit Risk, Profit and Loss Distribution, Density Optimization, Entropy Distribution, Probabilities of Default.

*The author would like to thank Faiyaz Bashar, Naser Bashir, Hans Beck, Jon Danielsson, Paul Embrechs, Charles Goodhart, Dennis Kristensen, Felix Muennich, Ryan Love, Pablo Padilla, Andrew Patton and Francisco Penaranda, for helpful discussions and useful comments. The author would also like to express gratitude to Global Asset Management (GAM) for their generous support and the motivation that was provided when they awarded this paper the first winner's prize for the GAM Gilbert de Botton Award in Finance Research 2004. The views expressed and any mistakes remain the author's. Those views do not necessarily reflect the views of the IMF, the CNBV or GAM.

Contents

List of Tables

List of Figures

1 Introduction

The credit risk of a bank's portfolio of loans is reflected in its profit and loss distribution (PLD). This distribution shows the possible gains and losses in the value of the portfolio and the related likelihood of such events, which are of prime importance for banks' economic capital decisions and their risk management strategies. Therefore, the proper measurement of the PLD has become a key objective in financial risk management. The PLD is a function of the changes in the credit risk quality of the loans that make up the portfolio. The credit risk quality of such loans depends on the value of the assets of the borrowing firms. Thus, modelling the PLD requires modelling the marginal and portfolio multivariate distributions that describe the individual and joint likelihood of changes in the value of the assets of the borrowing firms; this is equivalent to the marginal and portfolio multivariate distributions that describe the individual and joint likelihood of changes in the credit risk quality of the loans that make up the portfolio.[1]

Although, in recent years credit risk measurement has been improving rapidly, especially for market-traded instruments, information constraints still impose severe limitations when trying to measure portfolio credit risk. Information restrictions and thus, modelling limitations exist because either credit risk modellers have arm's-length relationships with the firms, i.e. arm's-length firms - and as a consequence those modellers[2] do not have access to the market and/or financial information that is necessary for the firms' risk assessment - or simply because there are certain variables that do not exist for the type of firms in which modellers are interested. The latter is the case when the modelling interest lies in the credit risk of loans granted to small and medium size enterprises (SMEs) and unlisted firms, i.e. closely held firms. Portfolio credit risk modelling under any of these restrictions constitutes the focus of our research interest.

The approaches that have been developed for the measurement of portfolio credit risk, the reduced form approach and the structural approach,[3] rely on assumed parametric distributions that, for proper calibration, need data that are not available to modellers or non-existent to closely held firms for reasons stated above.

The structural approach is one of the most common approaches to the modelling of loan portfolio credit risk, and therefore for modelling the PLD.[4] Under this approach, the change in the credit risk quality of a firm is a function of changes in the value of its assets. The basic premise of this approach is that a firm's underlying asset value evolves stochastically over time and default is triggered (i.e. the firm falls in the default state) by a drop in the firm's asset value below a threshold value, the latter being modelled as a function of the firm's financial structure. In its most basic version (Merton, 1974), the structural approach assumes that the firm's logarithmic asset returns are normally distributed; thus, the credit risk of a portfolio is described by a multivariate normal distribution with a dependence structure that is usually fixed through time. However, it has been empirically observed that both the normal distribution and the fixed dependence structure assumptions are not appropriate for financial instruments; therefore, risk managers and regulators have proposed different parametric and dependence modelling assumptions in order to improve the measurement of portfolio credit risk under the structural approach. Among these, we can find multivariate t-distributions with marginals containing the same or different degrees of freedom, historical simulation, mixture models and, more recently, copula functions.[5]

For any of these parametric assumptions, the availability of variables, or proxies, indicating the evolution of the firm's underlying asset value is of crucial importance for their proper calibration and implementation.

[1] Note that in this paper we make the assumption that the credit risk quality of a loan is the same as the credit risk quality of the borrowing firm to whom the loan is granted (debtor). Sometimes assets used as collateral can make the credit risk quality of the loan differ from the credit risk quality of the borrowing firm.

[2] Those modellers are usually financial regulators, or any authority that does not have direct relationships with the analysed firms, for example, IMF economists trying to assess credit risk in Financial Stability Assessment Programs (FSAPs).

[3] Reduced form models have been used to model the behavior of credit spreads. This approach treats default as a jump process with exogenous intensity (Duffie and Singleton, 1999). These models rely on variables that are only available for market-traded companies (e.g. bond yield spreads) to calibrate the intensity function for each obligor. Therefore, to our knowledge, reduced form models can only be applied to a small part of the fixed income market. Structural models are explained below.

[4] Widely known applications include the CreditMetrics framework (Gupton et al, 1997) and the KMV framework (Crosbie et al, 1998).

[5] Empirical support for modelling univariate returns with *t-like* tails can be found in Danielsson and de Vries (1997) and Hoskin et al (2000). Glasserman et al (2000) present the multivariate modeling framework. Historical simulation for portfolio credit risk is presented by Mina et al (2001) and Butler et al (1998). For mixture models see McLachlan et al (1987) and Zangari (1996).

Hence, any attempt to apply the structural approach - assuming the parametric distributions that have been proposed in its basic or improved versions - becomes a challenging prospect when focusing in closely held or arm's-length firms. In these restricted data environments, the frequency of loan defaults (PoDs) for a given classification of loans, e.g., loans classified under sectoral activities or credit risk qualities (ratings), is the only information that is usually available.

PoDs represent partial information on the marginal distribution of each type of loan within a portfolio. Nor, at the portfolio level, is it possible to observe the joint likelihood of changes in the asset value of the borrowing firms, nor variables indicating the joint likelihood of credit risk quality changes of the loans making up a portfolio. As a result, when modellers try to specify portfolio multivariate distributions from the incomplete set of information provided by the PoDs, they face an under-identified mathematical problem. When problems of under-identification arise, parametric assumptions, representing information that is not available, are typically used to convert under-identified problems into well-identified problems. However, this course of action is not an appropriate framework when attempting to model portfolio credit risk under the information restrictions that constitute the focus of our research interest. The lack of data makes it impossible to adequately calibrate the assumed parametric distributions; thus, these distributions may not be consistent with the analyzed assets' data-generating processes. As a consequence, erroneous statistical inferences and incorrect economic interpretations may result.[6]

Therefore, rather than imposing convenient distributional assumptions, this paper proposes an alternative route to recover portfolio multivariate distributions from the incomplete set of information available for the modelling of the portfolio credit risk of loans granted to SMEs, unlisted and arm's-length firms. We refer to this procedure as the Consistent Information Multivariate Density Optimizing methodology (CIMDO). CIMDO is based on the Kullback (1959) cross-entropy approach. This approach reverses the process of modelling data information. Instead of assuming parametric probabilities to characterize the information contained in the data, the entropy approach uses the data information to infer values for the unknown probability density. Because the information requirements necessary for the implementation of CIMDO are less stringent than those necessary for the proper calibration of parametric distributions, the implementation of CIMDO becomes straightforward.[7] CIMDO also seems to reduce the risk of density misspecification because it recovers densities that are consistent with empirical observations. This is evident when we present a simulation exercise, which provides evidence showing that CIMDO-recovered distributions outperform widely used distributions in portfolio credit risk modelling under the Probability Integral Transformation (PIT) criterion.

The implementation of CIMDO has the potential for wider relevance. In some countries, SMEs and unlisted firms represent the backbone of the economy, making a significant contribution to their GDP and to the sustainability of their employment levels. Moreover, loans granted to SMEs and unlisted companies usually represent an important percentage of the assets held by banks in most developed and developing economies.[8] Improvements in the methodologies used to measure the portfolio credit risk of these types of financial assets can have important implications for individual banks' risk management and for a system's financial stability.[9] The latter can be further enhanced by the ability of regulators - or any other authority with arm's-length relationships with the firms - to improve the measurement of credit risk in the system.

The remainder of this paper is structured as follows. In Section 2, we develop the CIMDO methodology. We provide its motivation by describing the structural approach (SA) to modelling credit risk. The description of the structural approach is not only useful to depict the intuition behind CIMDO. It is also helpful to shed light on the limitations embedded in the information that is available for the modelling of credit risk of loans granted to closely held and arm's-length firms. We continue by presenting the major results behind the principle of minimum cross-entropy, which is the theoretical framework on which CIMDO is

[6] Koyluoglu (2003) presents an interesting analysis of the consequences of the improper calibration of credit risk models.

[7] The empirical frequency of default (PoD) of each type of loan making up a portfolio is the only information necessary to recover CIMDO multivariate distributions. Information of the dependence structure of a multivariate distribution is not necessary to recover it. However, if available, such information can easily be incorporated into the modelling framework.

[8] For example, Saurina and Trucharte (2004) report that, in Spain, exposures to SMEs represent, on average, 71.4% of total bank exposures to firms. Similar results have been reported in the case of Germany. The Comision Nacional Bancaria y de Valores (Financial Regulatory Agency in Mexico) reports that the percentage of this type of firm is about 85% in Mexico.

[9] While we restrict our attention to loans, the proposed procedure can easily be extended to measure the portfolio credit risk of baskets, credit derivatives, mortgage backed securities, or any other synthetic instrument that holds underlying assets with similar data constraints to those of closely-held and arm's-length firms.

built. We then provide the CIMDO modelling framework and present a detailed mathematical development of the methodology. We end the section by discussing the CIMDO-inferred distribution. Section 3 presents the theoretical background and proofs behind the PIT, the criterion selected to perform density evaluations. In Section 4 we explore the robustness of CIMDO-recovered distributions. Section 5 presents an empirical implementation exercise. This section describes the data employed in this exercise and shows in detail how CIMDO is implemented to recover a portfolio distribution. It also shows how credit risk modellers would usually calibrate parametric portfolio distributions when faced with a similar information set. Density evaluations are performed and results are presented. Section 6 discusses the results obtained in the previous Section. Section 7 concludes.

2 Consistent Information Multivariate Density Optimizing Methodology

2.1 CIMDO: motivation

The structural approach is one of the most common approaches to the modelling of loan portfolio credit risk and loan portfolios' profit and loss distributions.[10] The CreditMetrics methodology (Gupton, Finger and Bhatia, 1997) and Moody's-KMV methodology (Crosbie, 1998) are widely known applications of this approach.

The basic premise of the structural approach is that a firm's underlying asset value evolves stochastically over time and default is triggered by a drop in the firm's asset value below a pre-specified barrier, henceforth called the default-threshold, which is modelled as a function of the firm's leverage structure. Therefore, once the parametric distribution driving the firm's underlying asset value and the default-threshold are defined, the firm's probability of default, which indicates the probability of the firm's asset value falling below the default-threshold, can be calculated.

The consistent information multivariate density optimizing methodology assumes that the basic premise of the structural approach holds; however, CIMDO reverses this process. Following a cross-entropy decision rule, from the observed probabilities of default of the loans making up a portfolio, CIMDO recovers the multivariate distribution followed by the underlying asset value of the firms making up the portfolio. In order to formalize these ideas, we find it useful to describe the structural approach.

The structural approach

In its most basic version (Merton,1974), the structural approach assumes that the firms' logarithmic asset values are normally distributed. Under this version of the structural approach, if it is assumed that a loan portfolio is composed of two types of borrowers with different credit risk qualities, we can describe the processes driving the borrowing firms' asset values as, $dS_t^x = \mu^x S_t^x dt + S_t^x \sigma^x dW_t^x$ and $dS_t^y = \mu^y S_t^y dt + S_t^y \sigma^y dW_t^y$, where $ln\left[S_t^x\right]$ and $ln\left[S_t^y\right]$ are normally distributed and W_t^x and W_t^y are Brownian motions with dependence structure $dW_t^x dW_t^y = \rho dt$. If it is also assumed that the initial logarithmic asset values are $ln\left[S_t^x\right] = 0$, $ln\left[S_t^y\right] = 0$, then $ln\left[S_T^x\right] \sim N\left[\left(\mu^x - \frac{1}{2}{\sigma^x}^2\right)(T-t),\, \sigma^x\sqrt{T-t}\right]$ and $ln\left[S_T^y\right] \sim N\left[\left(\mu^y - \frac{1}{2}{\sigma^y}^2\right)(T-t),\, \sigma^y\sqrt{T-t}\right]$. Therefore, we can represent the standardized logarithmic asset values of these borrowers at time T, as $x(T) = \frac{ln\left[S_T^x\right] - \left(\mu^x - \frac{1}{2}{\sigma^x}^2\right)(T-t)}{\sigma^x\sqrt{T-t}}$ and $y(T) = \frac{ln\left[S_T^y\right] - \left(\mu^y - \frac{1}{2}{\sigma^y}^2\right)(T-t)}{\sigma^y\sqrt{T-t}}$. As a result, $x(T) \sim \Phi(0,1)$ and $y(T) \sim \Phi(0,1)$. Then each borrower defaults at some time $T > t$, if the firm's value falls below its default-threshold X_d^x and X_d^y, i.e. $x(T) \leq X_d^x$, $y(T) \leq X_d^y$. Therefore, at time t, the firm's marginal probabilities of default, can be represented by

$$\begin{aligned} PoD_t^x &= \Phi\left(X_d^x\right), \\ PoD_t^y &= \Phi\left(X_d^y\right), \end{aligned} \tag{1}$$

[10] In order to generate the PLD using the structural approach, first, a portfolio multivariate distribution is specified; second, Monte Carlo simulation algorithms are used to generate simulated asset values from the specified distribution. With the simulated values, the PLD is estimated.

where $\Phi(\cdot)$ is the standard normal cumulative distribution function (*cdf*).

Consequently, the joint default probability of these borrowers is

$$PoD_t^{xy} = \Phi_\rho \left(X_d^x, X_d^y\right), \tag{2}$$

where $\Phi_\rho(\cdot)$ stands for the bivariate standard normal distribution function with a *time independent instantaneous* correlation structure ρ.

In this case, due to the assumption of the correlation structure between the Brownian motions, the dependence structure between the firms' asset values is described by[11]

$$\begin{aligned} Corr\left(dS_t^x, dS_t^y\right) &= \frac{Cov\left(dS_t^x, dS_t^y\right)}{\sqrt{Var\left(dS_t^x\right)}\sqrt{Var\left(dS_t^y\right)}} \\ &= \frac{S_t^x \sigma^x S_t^y \sigma^y dW_t^x dW_t^y}{\sqrt{Var\left(dS_t^x\right)}\sqrt{Var\left(dS_t^y\right)}} = \frac{S_t^x \sigma^x S_t^y \sigma^y \rho dt}{S_t^x \sigma^x \sqrt{dt} S_t^y \sigma^y \sqrt{dt}} \\ &= \rho. \end{aligned} \tag{3}$$

More recently, different parametric distributions have been proposed to improve the modelling of the stochastic behavior of the asset values in the structural approach. Among these, those which are most commonly used to measure loan credit risk are the t-distribution and mixture of normals. These distributions have been justified by the fact that financial assets' returns exhibit heavier tails than would be predicted by the normal distribution. It is common to observe large events, e.g. extreme asset price movements, more frequently than the normal distribution predicts.

Irrespective of the parametric assumptions that the modeler decides to take, the availability of variables, or proxies, indicating the evolution of the firms' underlying asset value, is of crucial importance for the proper specification of the assumed distributions.

Thus, the implementation of the structural approach to model the credit risk of SMEs and unlisted firms is a difficult task, since there is usually only very limited information to model the portfolio credit risk of these firms.[12] In the majority of cases, the empirical frequencies of default (PoD) for specific loan groups, e.g., loan credit-risk classes (ratings) or loans aggregated by sectors, are the only information that is available.

[11] For listed firms, it has become common practice to use bond yield spreads or proxies like equity returns and equity return correlations as proxies for asset returns and asset returns correlations. This is because, even for listed firms, the asset return and asset return correlations are not observed in practice. Theoretically, it can be proved that the local correlation between share prices should be equal to the local correlation between the underlying firm's value processes. Unfortunately, empirically there is mixed evidence about this theoretical result. While common practice in the industry is to use equity return correlations as proxies for asset return correlations, Servigny and Renault (2002) and Schonbucher (2003) document the fact that equity return correlations tend to be significantly higher than asset return correlations. There is no resolution to this contradiction, whatever explanation is preferred; the fact remains that caution has to be exercised when using equity correlations as a proxy for asset correlations. Nonetheless, although the lack of data is in general a significant problem for the measurement of credit risk, for listed firms, in some cases, there might be sufficient information available to make reasonable calibrations of the parametric distributional assumptions that are usually taken.

[12] For bank regulators, modelers who do not have direct relationships with the firms or secondary banks who did not originate the loans, restrictions on information are even tougher, since they lack access to proprietary data on the financial structure of these borrowers.

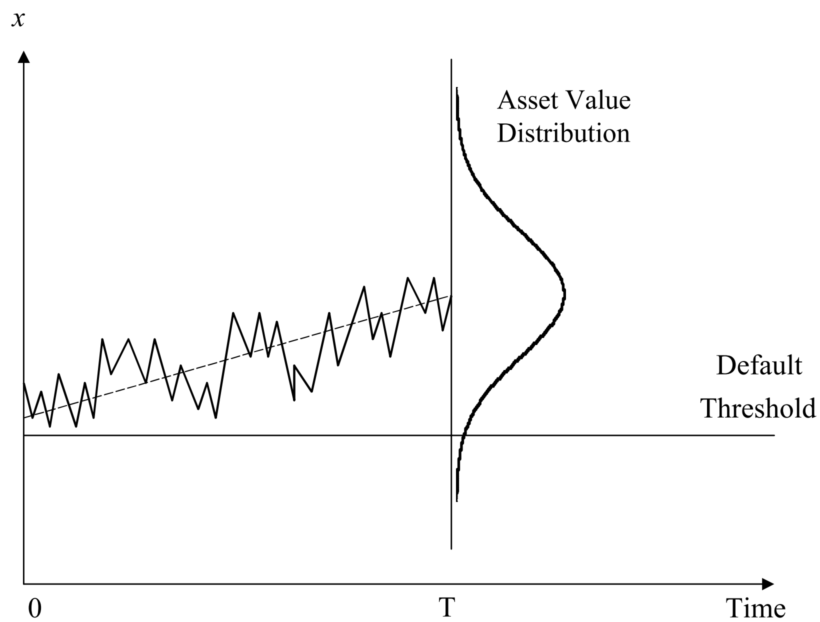

Figure 1: The structural approach

Empirical frequencies of default

Loans granted to SMEs and unlisted firms are usually classified and aggregated according to their sectoral activity or credit risk quality (rating).[13] For loans classified under a given sector or rating[14], the frequency of loan defaults, henceforth referred as the empirical probability of default (PoD) at different points in time, usually represents the only statistic that is available for the modelling of credit risk.[15] Time series of PoDs are also restricted through time, since, commonly, it is only possible to obtain very few observations with quarterly or yearly frequencies.

At a given period of time t, once a loan is grouped under a given classification m, the probability of default for that group is characterized as

$$PoD_t^m = \frac{n_{tdef}^m}{n_t^m}, \tag{4}$$

where n_t^m represents the total number of firms that are classified under the category m and n_{tdef}^m represents the number of firms that are classified under the category m that defaulted on their loan obligations during period t. Most commonly, the proportion PoD_t^m is the variable that is reported; therefore, not even information on the numerator or denominator is usually known.

[13] Loans granted to SMEs and unlisted firms are sometimes classified and aggregated according to other loan characteristics. These include the geographical location of the firms holding them, type of collateral of the loan etc. It is important to point out that our methodology can easily be applied under any of these types of aggregations.

[14] Credit risk quality classification statistics (ratings) classify loans according to their credit default risk. Ratings are usually produced by public rating agencies or by internal rating models. The most well known rating agencies include Standard and Poor's, Moody's and Fitch. Most ratings given to SMEs and unlisted companies are produced by internal rating models. The latter include models à la Altman (1989) or different applications based on the structural approach. For example, Moody's-KMV has derived a classification statistic known as the *distance to default* (DD). The DD measures the standardized distance (i.e. by the asset standard deviation) of a firm's value from its default-threshold. Thus, the larger the DD, the better the credit risk quality of a borrower. This statistic is assumed to contain all the relevant default information of the borrower. Once the DD have been obtained, they are used to classify firms that have similar DDs, as firms with a similar credit risk quality.

[15] For loans granted to SMEs and unlisted firms, it is possible, although only very rarely, to obtain frequencies of credit risk quality changes (rating transitions). However, for agents with *arms-length* relationships with the firms, it is difficult to verify the quality of these statistics, except for the default state, which is a state that cannot be "disguised" or hidden.

Why is information embedded in the PoDs not adequate for the proper calibration of parametric assumptions?

The PoDs represent merely *indirect* and *partial information* on the borrowing firms' asset value distributions. In order to make this claim clearer, first we assume that the structural approach holds and that the probability of default is characterized in the upper part of the distribution. Then, instead of assuming the normality assumption, as in equation (1) and (2), we allow the stochastic processes driving the borrowers' asset values to be represented by any multivariate distribution. Therefore, we can generalize the representation of the borrowers' marginal default probabilities as[16]

$$\begin{aligned} \int_{-\infty}^{\infty}\int_{X_d^x}^{\infty} p(x,y)dxdy &= PoD_t^x, \\ \int_{-\infty}^{\infty}\int_{X_d^y}^{\infty} p(x,y)dydx &= PoD_t^y, \end{aligned} \tag{5}$$

and, equally, we can generalize the representation of the borrowers' joint default probability as

$$\int_{X_d^y}^{\infty}\int_{X_d^x}^{\infty} p(x,y)dxdy = PoD_t^{xy}, \tag{6}$$

where X_d^x, X_d^y represent the default thresholds, $p(x,y)$ represents the portfolio multivariate distribution driving the borrowers' asset values, PoD_t^x, PoD_t^y represent the marginal empirical probabilities of default and PoD_t^{xy} represents the empirical joint default probability.

Note that the information provided by the empirical marginal probabilities of default, as described in equation (4), represents only *indirect information* on the portfolio marginal distributions, and as a result on the portfolio multivariate distribution, as can be appreciated in equation (5).

Moreover, the PoDs only provide information on the *region of default* of the marginal distributions of each type of loan in the portfolio. The latter can be represented by the interval of the borrower's asset value distribution, where the default state is triggered. If we set the region of default for each obligor in the upper part of a distribution, the regions of default for each borrower in the portfolio can be formalized mathematically with an indicating function as follows:

$$\chi_{[X^x d,\infty)} = \begin{cases} 1 & if \quad x \geqslant X_d^x \\ 0 & if \quad x < X_d^x \end{cases} \text{ and } \chi_{[X^y d,\infty)} = \begin{cases} 1 & if \quad y \geqslant X_d^y \\ 0 & if \quad y < X_d^y \end{cases}, \tag{7}$$

where the indicating function $\chi_{[X^x d,\infty)}$ and $\chi_{[X^y d,\infty)}$take the value of one in the region of default of each borrower in the portfolio, X_d^x and X_d^y represent the default-threshold values that delimit their regions of default. The region of default for an specific type of borrower can be appreciated in Figure 2 as the region where $x \geqslant X_d$ under an assumed distribution.

Equivalently, if one assumes that the basic premise of the structural approach holds, the marginal probability of default for each type of borrower in the portfolio only provides information on the frequency of firms whose asset value has crossed the default-threshold. However, apart from knowing that a given frequency of firms had asset values (the random variable in the model) that crossed the default threshold and moved into the default interval of the distributions' domain, we do not have any information on the specific values taken by the random variable nor the probabilities of the specific values of the random variable, neither in the region of default nor in any other region of the probability distribution's domain. Even worse, at the portfolio level, it is not possible to observe any variable representing the joint likelihood of asset value changes of the borrowers whose loans make up a portfolio, nor even indirect observations representing the joint likelihood of default of such borrowers.

Therefore, if from the incomplete and indirect information provided by the marginal empirical probabilities of default one tries to define the specification of the portfolio multivariate distribution $p(x,y)$ the problem is *under-identified,* since the possible specifications of $p(x,y)$ that satisfy equation (5) are infinite. Thus, when under-identified problems arise, it is necessary to define a basis for selecting a particular solution.

[16] In this representation, we assume that the portfolio consists of two types of loans.

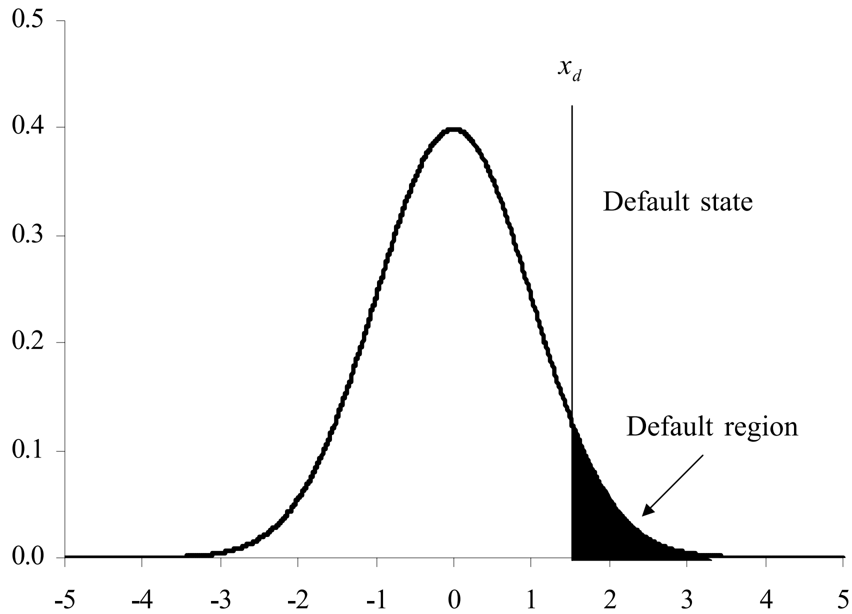

Figure 2: The region of default

The traditional route is to impose parametric distributional assumptions, representing information that we do not have. As we have already explained, the most common parametric distributions assumed under the SA are the Conditional Normal distribution, the t-distribution and the mixture of normals.[17]

However as we claimed above, for SMEs and unlisted firms, parametric assumptions do not provide an appropriate framework since it is not possible to observe variables or proxies indicating the evolution of the underlying asset value of these firms. Nor, at the portfolio level, is it possible to observe the joint likelihood of credit risk quality changes; nor, to observe variables indicating the dependence structure of the loans making up a portfolio.

As a result, it becomes impossible to calibrate adequately the assumed parametric distributions, as shown in Section 5; thus, these distributions may not be consistent with the analyzed assets' data-generating processes.

Consequently, rather than imposing parametric distributional assumptions, we follow an alternative route, using the entropy selection criteria embedded in CIMDO.

2.2 CIMDO: modelling foundation, the cross entropy approach

In order to recover a density $p(x)$, we propose a procedure based on the Kullback (1959) cross-entropy approach. When under-identified problems arise, this approach provides a rationale for selecting a solution out of the infinite possible ones. This procedure reverses the process of modelling data information. Rather than assuming parametric density functions, the cross-entropy approach uses the available information to infer the unknown probability density.

In order to formalize these ideas, we start with the exposition of the "entropy of the distribution of probabilities", an axiomatic method used by Shannon (1948) to define a unique function to measure the uncertainty of a collection of events.[18]

[17] We refer to the term "Conditional" because as explained in Section 3.4.3, the most common assumption made by risk modellers is that firms' logarithmic asset values follow a normal or t-distribution conditioned on the period's level of volatility.

[18] The origins of the entropy concept go back to the *XIXth* century with the work developed by Boltzman and continued subsequently by Maxwell, Gibbs, Bernoulli and Laplace.

In developing this approach, Shannon (1948) supposed that an experiment with N trials (repetitions) was carried out. This experiment had K possible outcomes (states). He assumed that $N_1, N_2, ..., N_K$ represented the number of times that each outcome occurs in the experiment of length N, where $\sum N_k = N$, $N_k \geqslant 0$, and $k = 1, 2, ..., K$.

In this setting there are N trials and each trial has K possible outcomes; therefore, there are K^N conceivable outcomes in the sequence of N trials. Of these, a particular set of frequencies

$$p_k = \frac{Nk}{N} \qquad \text{or} \qquad N_k = Np_k \qquad \text{for} \qquad k = 1, 2, ..., K,$$

can be realized in a given number of ways as measured by the multiplicity factor (possible permutations). Thus, the number of ways that a particular set of N_k is realized, can be represented by the multinomial coefficient

$$W = \frac{N!}{Np_1!Np_2!....Np_k!} = \frac{N!}{\prod_k N_k!},$$

or its monotonic function

$$lnW = lnN! - \sum_{k=1}^{K} \ln N_k! \tag{8}$$

Shannon (1948) manipulated equation (8), as presented in Appendix A.1, to obtain the Shannon (1948) entropy measure, which is defined as

$$E(p_k) = -\sum_k p_k \ln p_k, \tag{9}$$

where X is a random variable with possible outcome values x_k, $k = 1, 2, ..., K$ and probabilities p_k such that $\sum\limits_k p_k = 1$ and where $p_k \ln p_k = 0$ for $p_k = 0$.

Jaynes (1957) proposed to make use of the entropy concept to choose an unknown distribution of probabilities when only partial information is available. He proposed to maximize the function presented in equation (9), subject to the limited available data, in order to obtain the probability vector $\mathbf{p}$ that can be realized in the greatest number of ways consistent with the known data.

The rationale provided by Jaynes (1957) for choosing a particular solution, i.e. probability vector $\mathbf{p}$ from partial information, is known as the principle of Maximum Entropy (MED). Let

$$L = -\sum_k^K p_k \ln p_k + \sum_{t=1}^{T} \lambda_t \left[y_t - \sum_k^K p_k f_t(x_k) \right] + \mu \left[1 - \sum_k^K p_k \right], \tag{10}$$

be the Lagrange function. Then, the problem of Maximum Entropy is to maximize L.

In this function, the information contained in the data has been formalized in $1 \leq t \leq T$ moment-consistency constraints of the form $\sum\limits_{k=1}^{K} p_k f_t(x_k) = y_t$. These moment-consistency-constraints are formulated with T functions $\{f_1(x), f_2(x),..., f_T(x)\}$ representing the information contained in the data and with a set of observations (averages or aggregates) $\{y_1(x), y_2(x),...,y_T(x)\}$ that are consistent with the distribution of probabilities $\{p_1, p_2,...,p_k\}$. Note that the problem is *under-identified* if $T < K$.

In this function, the additivity restriction $\sum\limits_{k=1}^{K} p_k = 1$ has to be fulfilled as well, since $\mathbf{p}$ represents a probability distribution. Note also that λ_t represents the Lagrange multiplier of each of the $1 \leq t \leq T$ moment-consistency constraints and μ represents the Lagrange multiplier of the probability additivity constraint.

Using the method of Lagrange multipliers, the Maximum Entropy solution is given by

$$\widehat{p}_k = \frac{1}{\sum\limits_{k=1}^{K} \exp\left[-\sum\limits_{t=1}^{T} \widehat{\lambda}_t f_t(x_k) \right]} \exp\left[-\sum_{t=1}^{T} \widehat{\lambda}_t f_t(x_k) \right]. \tag{11}$$

An extension to the rationale provided by Jaynes is the Minimum Cross Entropy Distribution (MXED) developed by Kullback (1959) and Good (1963). Under the MXED, it is assumed that, in addition to the T moment constraints, some form of conceptual knowledge exists about the properties of the system and that this knowledge can be expressed in the form of a *prior* probability vector $\mathbf{q}$. In contrast to the MED pure inverse problem framework, the objective may be reformulated as being to minimize the *cross entropy distance* between the *posterior* $\mathbf{p}$ and the *prior* $\mathbf{q}$.[19] The cross-entropy objective function is defined as follows:

$$C\left[p_k, q_k\right] = \sum_{k}^{K} p_k \ln\left[\frac{p_k}{q_k}\right], \tag{12}$$

subject to T moment-consistency constraints $\sum_{k=1}^{K} p_k f_t(x_k) = y_t$, and the additivity restriction $\sum_{k=1}^{K} p_k = 1$. Consequently, the probability vector $\mathbf{p}$ may be recovered by minimizing the Lagrangian function

$$L = \sum_{k}^{K} p_k \ln\left[\frac{p_k}{q_k}\right] + \sum_{t=1}^{T} \lambda_t \left[y_t - \sum_{k}^{K} p_k f_t(x_k)\right] + \mu\left[1 - \sum_{k}^{K} p_k\right], \tag{13}$$

where λ_t and μ represent Lagrange multipliers. Using the method of Lagrange multipliers, the optimal solution to the cross-entropy problem is

$$\widehat{p}_k = \frac{q_k}{\sum_{k=1}^{K} q_k \exp\left[\sum_{t=1}^{T} \widehat{\lambda}_t f_t(x_k)\right]} \exp\left[\sum_{t=1}^{T} \widehat{\lambda}_t f_t(x_k)\right]. \tag{14}$$

Shannon's entropy measure $E(p_k)$, presented in equation (9), can be interpreted as a measure of the degree of uncertainty in the distribution of probabilities, representing the extent to which the distribution is concentrated on specific points as opposed to being dispersed over many points. $E(p_k)$ takes its maximum value lnK, when the distribution is maximally dispersed and thus uniformly distributed on the range of the distribution, in other words, when $p_1 = p_2 = ... = p_k = \frac{1}{K}$. $E(p_k)$ takes its minimum value 0, when the distribution is maximally informative in that p degenerates on one particular x_k value, in other words, $p_k = 1$ and $p_j = 0$, $\forall k \neq j$.

The objective of maximizing $E(p_k)$ in the absence of any constraints, other than the additivity restriction $\sum_{k=1}^{K} p_k = 1$, can be interpreted as choosing p_k to be the maximally uniform distribution. This is because, as mentioned above, $E(p_k)$ takes its maximum value when $p_1 = p_2 = ... = p_k = \frac{1}{K}$.

However, in the presence of T moment constraints, as suggested by Jaynes, the objective can be interpreted as choosing the p_k*'s* to be as maximally uninformative *as the moment constraints allow.* This objective is consistent with the goal of not wanting to assert more of the distribution p_k, than is known. In other words, irrelevant information is *"minimized out"*.

When, in addition to the T moment constraints, supplementary information in the form of a *prior* probability is incorporated into the optimization framework, the Kullback cross entropy framework recovers the density that could have been generated in the greatest number of ways consistent with these constraints and that has the smallest entropic distance from the *prior* distribution.

The *entropic distance* between p_k and q_k *is not* a metric distance[20] but it does satisfy $C\left[p_k, q_k\right] = 0$ for $p_k = q_k$ and $C\left[p_k, q_k\right] > 0$, whenever $p_k \neq q_k$. Note that the definition of $C\left[p_k, q_k\right]$ does not contain a negative sign, so cross entropy is minimized rather than maximized. Note also that equation (14) reduces to equation (11) when $q(x)$ is constant, i.e. a uniform distribution indicating no prior information about X.

[19] See Appendix 2.A.3 for the meaning of the *prior* $\mathbf{q}$.

[20] This is because $C\left[p(x), q(x)\right] \neq C\left[q(x), p(x)\right]$, however, for our objective, this property is not particularly important because the *prior* distribution is taken to be fixed in estimation problems. Our interest centers on choosing the *posterior* distribution that solves the moment equations and that is as close as possible to the *prior* distribution.

2.3 CIMDO: modelling framework

Based on the Kullback (1959) cross-entropy approach, CIMDO provides a rationale to infer the unknown multivariate distribution of a loan portfolio containing M different credit-risk-quality (rating) categories of loans, $p(l^1, .., l^M) \in \mathbb{R}^M$, henceforth referred to as the *posterior distribution*, from the observed PoDs of the loans making up the portfolio.[21]

For this purpose, under *an initial hypothesis*, it is assumed that the multivariate distribution driving the stochastic process of the portfolio follows a parametric distribution $q(l^1, .., l^M) \in R^M$, henceforth referred to as the *prior distribution*. It is important to point out that the initial hypothesis is taken in accordance with economic intuition (default is triggered by a drop in the firm's asset value below a threshold value) and with theoretical models (structural approach) but not necessarily with empirical observations. Thus, constraint equations are formulated with the information provided by the empirically observed PoDs of the loans making the portfolio.

Lastly, following the minimum-cross entropy principle embedded in CIMDO, the posterior density is recovered. This is t*he distribution that is closest to the prior distribution but that is consistent with the empirically observed probabilities of default* of the loans making up the portfolio.

Mathematically, the problem is converted from one of deductive mathematics to one of inference involving an optimization procedure. In order to formalize these ideas, we proceed by defining the objective function and the moment-consistency constraints that are necessary to set up the optimization procedure used by CIMDO.

2.3.1 Objective function

For a loan portfolio, containing loans given to M different classes of borrowers, whose logarithmic returns are characterized by the random variables $l^1, .., l^M$, we define the objective function to be used as

$$C\,[p, q] = \int\limits_{l^M} .. \int\limits_{l^1} p(l^1, .., l^M) \ln \left[\frac{p(l^1, .., l^M)}{q(l^1, .., l^M)} \right] dl^1..dl^M,$$

where $q(l^1, .., l^M)$ the prior distribution and $p(l^1, .., l^M)$ the posterior distribution $\in \mathbb{R}^M$.

In the interest of parsimony, from now on, we develop the multivariate case as a bivariate problem, although, all the results presented for the bivariate case are directly applicable for $\mathbb{R}^M$ when $M \geq 2$.

We focus on a portfolio containing loans given to two different classes of borrowers, whose logarithmic returns are characterized by the random variables x and y, where $x, y \in l^i$ s.t. $i = 1, .., M$. Therefore, the objective function can now be defined as:

$$C\,[p, q] = \int \int p(x, y) \ln \left[\frac{p(x, y)}{q(x, y)} \right] dxdy, \tag{15}$$

where $q(x, y)$ the prior distribution and $p(x, y)$ the posterior distribution $\in \mathbb{R}^2$.[22]

The intuition behind this objective function is heuristically discussed in Section 2.5.

The prior distribution

As already mentioned, CIMDO assumes that a default is triggered by a drop in the firm's asset value below a threshold value. Therefore, taking the structural approach as a departing point for our modelling, and assuming that its basic premise and economic intuition are correct, we set as an *initial hypothesis* that the portfolio follows a multivariate distribution, $q(x, y) \in \mathbb{R}^2$ that is Normal $N(0, I)$, where I is the identity matrix.

The justification for this parametric distribution is based on the parametric assumption behind the basic version of the structural approach (Merton,1974). The dependence assumption on this distribution comes from the fact that, as explained in Section 2.1, for the type of loans that are the subject of our

[21] CIMDO is built on the consistent information density optimising Methodology (CIDO), which recovers univariate densities. CIDO is presented in Chapter 2 of this thesis.

[22] This objective function is to be minimised in the argument $p(x, y)$ subject to a set of contraints, as indicated in equation (20).

interest, we only get partial information represented by the PoDs of each type of loan in the portfolio, and the time series of PoDs are usually very short, thus making it very difficult to define a specific dependence structure.

We emphasize that the initial hypothesis seems reasonable before to analyzing the information provided by the sample of loans included in the portfolio. Of course, after this information is considered, the *prior distribution* $q(x, y)$ may be found to be incompatible with the sample information.

2.3.2 Moment-consistency constraints

As we mentioned above, the prior distribution is chosen due to its consistency with theoretical arguments and economic intuition but not necessarily with empirical observations. Thus, the information provided by the frequencies of default of each type of loan making up the portfolio is of prime importance for the recovery of the posterior distribution. In order to incorporate this information into the recovered posterior density, we formulate the moment-consistency constraint equations.

Intuitively, we would like to update the shape of the portfolio multivariate distribution at each period of time, such that the *posterior* multivariate distribution is consistent with the empirically observed probabilities of default. In other words, the posterior multivariate distribution fulfils a set of restrictions imposed on its marginal distributions. These restrictions imply that, at each period of time, the region of default defined in equation (7) for each of the borrowers' marginals are equalized to each of the borrowers' empirically observed frequencies of default, which, at each period of time, change due to numerous factors affecting the underlying asset value of the firm, for example, macroeconomic shocks and variations in the business cycle.

Therefore, we fixed the threshold delimiting the region of default for each marginal. Once this threshold is fixed, we allow the shape of the posterior distribution to vary according to the empirical frequencies of default of each of the borrowers at each period of time.[23] Consequently, we characterize the default-threshold as follows.

Default-threshold

In order to fix the default-threshold, we define a "through-time-average default-threshold". For this purpose, for each type of borrower, we define the "through-time-average probability of default" as the historical average probability of default, $\overline{PoD}^m$.[24]

Given the assumed prior distribution, we characterize the "through-time-average default threshold" for each borrower as

$$X_d^x = \Phi^{-1}\left(\overline{\alpha}^x\right) \text{ and } X_d^y = \Phi^{-1}\left(\overline{\alpha}^y\right), \tag{16}$$

where we defined $\overline{\alpha}^x = 1 - \overline{PoD}^x$ and $\overline{\alpha}^x = 1 - \overline{PoD}^x$ since, in our model, the region of default for each obligor is described in the upper part of a distribution, and $\Phi(\cdot)$ is the standard normal *cdf*.

Hence, information on the empirical probabilities of default for each type of borrower allows us to find their particular default thresholds without the need of variables indicating the evolution of their underlying asset values and their financial structures.

[23] Equivalently, the optimization procedure embedded in CIMDO changes the shape of the *posterior* multivariate distribution in a way that ensures the satisfaction of the restrictions imposed on its marginals at each period of time.

[24] Depending of the way that PoD_t^m are reported, the $\overline{PoD}^m$ can be defined in different ways. Ideally, we would like to be able to define it as $\overline{PoD}^m = \frac{\sum_{t=1}^{T} n_{tdef}^m}{\sum_{t=1}^{T} n_t^m}$ where n_t^m represents the total number of firms classified under the rating category m and n_{tdef}^m is the number of firms classified under the rating category m that defaulted on their loan obligations during period t, and $t = \{1, 2, ...T\}$, represents the number of periods for which PoD_t^m are available. However, if information of the numerator and denominator is not available, $\overline{PoD}^m$ can be computed as a through-time average. In this case, modellers should be careful with the weights assigned to each observation.

Moment-consistency constraints

Once the default-thresholds are defined, we proceed to formalize the definition of the moment-consistency constraints as[25]

$$\begin{aligned} \int_{X_d^x}^{\infty} dP_x(x) &= PoD_t^x, \\ \int_{X_d^y}^{\infty} dP_y(y) &= PoD_t^y, \end{aligned} \tag{17}$$

where $P_x(x)$, $P_y(y)$ are the *posterior cdf's* and PoD_t^x, PoD_t^y are the empirically observed probabilities of default for each borrower in the portfolio. Equivalently, we define the constraints in terms of their *pdfs* as

$$\begin{aligned} \int p(x)\chi_{[X_d^x,\infty)}dx &= PoD_t^x, \\ \int p(y)\chi_{[X_d^y,\infty)}dy &= PoD_t^y, \end{aligned} \tag{18}$$

where $p(x)$, $p(y)$ are the *posterior pdfs*, $\chi_{[X_d^x,\infty)}$, $\chi_{[X_d^y,\infty)}$ are the indicating functions as defined in equation (7) and PoD_t^x, PoD_t^y are the empirically observed probabilities of default for each borrower in the portfolio.

Nonetheless, our objective is to recover the portfolio multivariate distribution. Therefore, the restrictions on the *pdfs* of each of the borrowers holding the loans making up the portfolio are expressed as restrictions on the marginals of the portfolio multivariate distribution,

$$\begin{aligned} \int_{-\infty}^{\infty}\int_{X_d^x}^{\infty} p(x,y)dxdy &= PoD_t^x, \\ \int_{-\infty}^{\infty}\int_{X_d^y}^{\infty} p(x,y)dydx &= PoD_t^y, \end{aligned}$$

or, equivalently, as

$$\begin{aligned} \int\int p(x,y)\chi_{[X_d^x,\infty)}dxdy &= PoD_t^x, \\ \int\int p(x,y)\chi_{[X_d^y,\infty)}dydx &= PoD_t^y, \end{aligned} \tag{19}$$

where $p(x,y)$ is the *posterior* multivariate distribution that represents the unknown to be solved.

Note that the moment-consistency constraints embed all the available information for each type of borrower. This information is included in the constraints via the default-threshold and the empirically observed probabilities of default at each period of time. Imposing these constraints on the optimization problem guarantees that the posterior multivariate distribution contains marginal densities that in the region of default, as defined in equation (7), are equalized to each of the borrowers' empirically observed probabilities of default.

[25] We refer to this constraint as the moment-consistency constraint, based on the definition of *partial moments*. Using definition 3.A.1.2 in Appendix 3.A.1, we set $Z = X_d^m$ and $r = 0$. Then, the probability that $l^m \geq X_d^m$ can be expressed as the partial moment $M\left(0, X_d^m\right) = \int_{X_d^m}^{\infty} dF_{l^m}$.

2.4 CIMDO: methodology set up

The Consistent Information Multivariate Density Optimizing methodology recovers the *posterior* multivariate density $p(x,y) \in \mathbb{R}^2$ by minimizing the objective function defined in equation (15) subject to the moment-consistency constraints defined in equation (19). In order to ensure that $p(x,y)$ represents a valid density, the conditions that, $p(x,y) \geqslant 0$ and the probability additivity constraint, $\int\int p(x,y)dxdy = 1$, also need to be satisfied.

The CIMDO density is recovered by minimizing the functional

$$\begin{aligned} L[p,q] &= \int\int p(x,y)\ln p(x,y)dxdy - \int\int p(x,y)\ln q(x,y)dxdy \\ &+\lambda_1\left[\int\int p(x,y)\chi_{[X_d^x,\infty)}dxdy - PoD_t^x\right] \\ &+\lambda_2\left[\int\int p(x,y)\chi_{[X_d^y,\infty)}dydx - PoD_t^y\right] \\ &+\mu\left[\int\int p(x,y)dxdy - 1\right], \end{aligned} \tag{20}$$

where, $p(x,y)$ is the *posterior* multivariate distribution, the unknown to be recovered, and $q(x,y)$ is the *prior* multivariate distribution as defined in equation (15). λ_1, λ_2 represent the Lagrange multipliers of the moment-consistency constraints defined in equation (19) and μ represents the Lagrange multiplier of the probability additivity constraint.[26]

This functional can be re-written as follows

$$\begin{aligned} L[p,q] &= \int\int p(x,y)\left[\ln p(x,y) - \ln q(x,y)\right]dxdy \\ &+\int\int p(x,y)\left[\lambda_1\chi_{[X_d^x,\infty)} + \lambda_2\chi_{[X_d^y,\infty)} + \mu\right]dxdy \\ &-\lambda_1 PoD_t^x - \lambda_2 PoD_t^y - \mu. \end{aligned} \tag{21}$$

By using the calculus of variations, the optimization procedure can be performed by computing the following variation,

$$\delta L = \frac{dL\left[p(x,y) + \epsilon\gamma(x,y), q(x,y)\right]}{d\epsilon}\Big|_{\epsilon=0} = 0,$$

where ϵ is a small quantity and $\gamma(x,y)$ is an arbitrary continuous function with a value of zero at the end points (boundary) of integration and with finite variance. Without loss of generality, we can assume that $\gamma(x,y)$ is bounded.

The optimal solution is represented by the following *posterior* multivariate density as

$$\widehat{p(x,y)} = q(x,y)\exp\left\{-\left[1+\widehat{\mu}+\left(\widehat{\lambda}_1\chi_{[X_d^x,\infty)}\right)+\left(\widehat{\lambda}_2\chi_{[X_d^y,\infty)}\right)\right]\right\}. \tag{22}$$

2.5 CIMDO: intuition

Analyzing the functional defined in equation (20), it is clear that the consistent multivariate density optimizing methodology recovers the distribution that minimizes the probabilistic divergence; i.e. "entropy distance", from the prior distribution and that is consistent with the information embedded in the moment-consistency constraints. Thus, out of all the distributions satisfying the moment-consistency constraints, the proposed procedure provides a rationale by which we select the posterior that is closest to the *prior*,

[26] Like all optimization problems, the Lagrange multipliers reflect the change in the objective function's value as a result of a marginal change in the constraint set.

thereby, solving the under-identified problem that was faced when trying to determine the unknown multivariate distribution from the partial information provided by the PoDs.

The intuition behind this optimization procedure can be understood by analyzing the cross-entropy objective function defined in equation (15). This objective function is an extension of the pure-entropy function presented in Section 2.2. The pure-entropy function described in equation (9) is a monotonic transformation of the multiplicity factor shown in equation (8). The multiplicity factor indicates the number of ways that a particular set of frequencies can be realized (i.e. this set of frequencies corresponds to the frequencies of occurrence attached to specific values of a random variable, Shannon, 1948). Therefore, when maximizing the entropy function subject to a set of constraints, we obtain the set of frequencies (frequency distribution) that can be realized in the greatest number of ways and that is consistent with the constraints (Jaynes, 1957). However, if an initial hypothesis of the process driving the behavior of the stochastic variable can be expressed in the form of a *prior distribution*, now, in contrast to the maximum entropy pure inverse problem, the problem can be reformulated to minimize the probabilistic divergence between the *posterior* and the *prior*. Out of all the distributions of probabilities satisfying the constraints, the solution is the *posterior* closest to the *prior* (Kullback, 1959). Although a *prior distribution* is based on economic intuition and theoretical models, it is usually inconsistent with empirical observations.[27] Thus, using the cross-entropy solution, one solves this inconsistency, reconciling it in the best possible way by recovering the distribution that is closest to the *prior* but consistent with empirical observations.

When we use CIMDO to solve for $p(x, y)$, the problem is converted from one of deductive mathematics to one of inference involving an optimization procedure. This is because the cross-entropy approach embedded in CIMDO reverses the process of modelling data information. Instead of assuming parametric probabilities to characterize information contained in the data, this approach uses the data information to infer values for the unknown probability density. The recovered probability values can be interpreted as inverse probabilities. Using this procedure, we look to make the best possible predictions from the scarce information that we have. This feature of the methodology not only makes implementation simple and straightforward, it also seems to reduce model and parameter risks of the recovered distribution, as indicated by the PIT criterion.[28] This is because in order to recover the *posterior* density, only variables that are directly observable for the type of loans that are the subject of our interest are needed. And because, once the *posterior* density is recovered, there is no need to calibrate it, since, by construction, the recovered *posterior* is consistent with the empirically observed probabilities of default. In this sense, the proposed methodology represents a more flexible approach to modelling multivariate densities, making use of the limited available information in a more efficient manner.

3 CIMDO: density evaluation criterion

The evaluation framework, used to assess whether CIMDO-recovered distributions provide improvements over the most commonly assumed parametric distributions for the measurement of loan portfolio credit risk is extremely important. In order to evaluate these densities, we follow the Probability Integral Transformation (PIT) approach developed by Diebold *et al* (1999). In what follows in this section, we provide detailed theoretical results supporting the chosen evaluation criterion.

The Probability Integral Transformation (PIT) Approach

Density evaluation is not a trivial problem, since there is no way to rank two incorrect density forecasts such that all users will agree with the ranking. Ranking depends on the specific loss functions of the users[29]. However, Diebold et al (1998) assert that "if a forecast coincides with a random variable true Data Generating Process (DGP), then it will be preferred by all forecast users, regardless of loss function". Thus this proposition implies that, regardless of the users' loss function, the correct density is weakly

[27] Note that if the information included in the *prior* q is consistent with the data (moment-consistency constraints), then $p = q$. This would be reflected in a multiplier with value equal to zero; i.e. the constraint is redundant and has no optimal informational value.

[28] We present evidence of the improvements of the distribution recovered by CIMDO in Section 5.

[29] The result is analogous to Arrow's celebrated impossibility theorem. The ranking effectively reflects a social welfare function, which does not exist.

superior to all forecasts. As a result, Diebold et al (1998) suggest *evaluating forecasts by assessing whether the forecast densities are correct*, though the task of determining whether a forecast equals the true DGP appears to be difficult because the true DGP is never observed. Moreover, the true density may exhibit structural changes at every point in time. To overcome this difficulty, they developed a method based on the Rosenblatt (1952) PIT approach.

Diebold et al (1998) prove that the Probability Integral Transformations of a series of density forecasts are distributed **iid** $\mathbf{U(0,1)}$ if the density forecasts coincide with the true DGP. *Thus, to assess whether a series of density forecasts coincides with the corresponding series of true conditional densities, it is only necessary to test whether the sequence of density forecast PITs are* **iid** $\mathbf{U(0,1)}$.

Diebold *et al* (1999) extend their results to the M_{th}-multivariate case. Suppose there is a series of T, M_{th}-multivariate density forecasts. They factorize each period's t, joint forecast density into the product of their conditionals:[30]

$$p_{t-1}(l_t^1, .., l_t^M) = p_{t-1}(l_t^M | l_t^{M-1}, ..l_t^1) .. \cdot p_{t-1}(l_t^2 | l_t^1) \cdot p_{t-1}(l_t^1). \tag{23}$$

This procedure produces a set of $(M-1)$ conditionals and (1) marginal density. The PITs of the l^m random variable realizations under these M series will be $\mathbf{iidU(0,1)}$, individually and also when taken as a whole, if the multivariate density forecasts are correct. The proof of this assertion in a time series framework can be found in Diebold et al (1999). The results of the univariate can be extended in a time series framework to show that there are M vectors (of dimension T) of PIT's that are **iid** $\mathbf{U(0,1)}$.

However, we need to develop a slightly different test. This is because CIMDO recovers densities with information at each period of time t, in other words, it updates period by period. Thus, the density evaluation needs to be done at specific periods of time t. Alternatively, rather than evaluating a time-series as Diebold et al (1999) do, we evaluate cross-sectionally the multivariate distribution. As a result, we need to prove that at each point in time t, the product of the conditionals and marginal PITs on which a multivariate distribution can be decomposed is **iid** $\mathbf{U(0,1)}$. This proof is not explicitly presented in Diebold et al (1999); therefore, we develop the proof for the bivariate case in what follows.

Recall that we focus on a portfolio containing loans of two different credit-risk-quality categories (ratings), whose logarithmic returns are characterized by the random variables x and y, where $x, y \in l^i$ s.t. $i = 1, .., M$.

The portfolio bivariate density can be decomposed into

$$p(x,y) = p(x) \cdot p(y|x), \tag{24}$$
$$p(x,y) = p(y) \cdot p(x|y). \tag{25}$$

We start by analyzing the first case presented in equation (24). Under the PIT, two new variables are defined as

$$\begin{aligned} u &= P(x) \Longleftrightarrow x = P^{(-1)}(u), \\ v &= P(y|x) \Longleftrightarrow y = P^{(-1)}(v|x). \end{aligned}$$

First we prove that u, v are distributed $\mathbf{U(0,1)}$, then we prove that they are independent.

Theorem 1 ***Probability Integral Transformation***

Let $\mathbf{x}$ be a random vector with absolute continuous *cdf* F. Define a new random variable, the "Probability Integral Transformation" (PIT) as

$$U = F(x). \tag{26}$$

Then, $U \sim \mathbf{U(0,1)}$ regardless of the original distribution F.[31]

[30] Note that the M_{th} multivariate density can be factorized into $M!$ ways at each period of time t.

[31] In this case, $U = F(x)$ is a generalization of $u = P(x)$ and $v = P(y/x)$.

Proof. $U \sim \mathbf{U(0,1)}$
For u on $[0,1]$, we have:
$$\begin{aligned} P[U \leq u] &= P[F(x) \leq u] \\ &= P\left[F^{-1}[F(x)] \leq F^{-1}(u)\right] \\ &= P\left[x \leq F^{-1}(u)\right] \\ &= F\left[F^{-1}(u)\right] \\ &= u \qquad \blacksquare \end{aligned}$$
For $u < 0$, $P[U < u] = 0$ and for $u > 1$, $P[U > u] = 0$ since the range of a *cdf* is $[0,1]$. Thus

Proposition. u, v are independent.
Proof.
In order to prove the independence assumption, we know that the joint density $c[u,v]$ is defined under the distribution of transformations of random variables as (Cassella and Berger, 1990)

$$c[u,v] = p\left[P^{(-1)}(u),\ P^{(-1)}(v|x)\right] \cdot \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}.$$

Since in this case

$$\begin{aligned} u &= P(x) \Longleftrightarrow x = P^{(-1)}(u) \Longrightarrow \frac{\partial x}{\partial u} = p\left[P^{(-1)}(u)\right]^{-1}, \\ v &= P(y|x) \Longleftrightarrow y = P^{(-1)}(v|x) \Longrightarrow \frac{\partial y}{\partial v} = p\left[P^{(-1)}(v|x)|x\right]^{-1}, \\ \frac{\partial x}{\partial v} &= \frac{\partial y}{\partial u} = 0, \end{aligned}$$

we get

$$\begin{aligned} c[u,v] &= p\left[P^{(-1)}(u),\ P^{(-1)}(v|x)\right] \cdot \frac{1}{p[x] \cdot p[y|x]}, \\ &= p[x,y] \cdot \frac{1}{p[x] \cdot p[y|x]}, \\ &= p(x) \cdot p(y|x) \cdot \frac{1}{p[x] \cdot p[y|x]}, \\ &= 1. \end{aligned} \tag{27}$$

which proves that u, v are independent. $\blacksquare$

In time series frameworks, empirical tests for **iid** $\mathbf{U(0,1)}$ are usually done in two stages. First independence is tested and then, conditional on the series being independent, uniformity is tested. In our particular case, independence is proved and therefore, it is not necessary to empirically test for it.

Finally, on the basis of these results, when we empirically compare different multivariate densities, we are able to determine that the specification of a given multivariate density will be better than alternative specifications, the closer the PITs of its marginals and conditionals are to **iid** $\mathbf{U(0,1)}$. Given the data limitations, we expect the "true distribution" to be unattainable, i.e., none of the compared distributions will be significantly close to $\mathbf{U(0,1)}$. Therefore, we focus on the relative improvements of the different distribution specifications.

4 CIMDO: robustness

We considered it interesting to explore the robustness of the CIMDO-recovered distributions with respect to two aspects.

1. Robustness with respect to changes (errors) in the optimal density, which represents the optimal solution to the CIMDO optimization framework.

2. Robustness with respect to perturbations in the assumed prior density.

In the interest of parsimony, we develop these proofs in a univariate setting, although all the results presented for the univariate case are directly applicable for $\mathbb{R}^M$ when $M \geq 2$.

These results offer an interesting perspective to compare the CIMDO-recovered density to other optimization frameworks.

4.1 Robustness w.r.t changes in the optimal solution

Because CIMDO is based on an optimization procedure, there is already an implicit robustness, i.e. stability, feature within the methodology. We formalize this concept mathematically by means of a Taylor series expansion.

We start this analysis by recalling that given a *prior* $q(x)$, and a *posterior* $p(x)$ the objective function can be defined by

$$C\left[p,q\right] = \int p(x) \ln \left[\frac{p(x)}{q(x)}\right] dx. \tag{28}$$

If, as is the case, there are additional restrictions on $p(x)$ of the form

$$R\left[p\right] \equiv \int f\left[p\left(x\right)\right] dx = ct, \tag{29}$$

the problem we are solving is to minimize $C\left[p,q\right]$ subject to $R\left[p\right]$.

Using the Lagrange multiplier method we know that the problem is equivalent to finding suitable critical points of

$$L\left[p,q\right] = C\left[p,q\right] + \lambda R\left[p\right]. \tag{30}$$

Substituting equation (28) and equation (29) into equation (30), we get

$$L\left[p,q\right] = \int \left\{p(x)\left[\ln p\left(x\right) - \ln q\left(x\right)\right] + \lambda R\left[p\left(x\right)\right]\right\} dx. \tag{31}$$

In order to analyze the Taylor series expansion, we compute the first variation of equation (31) for a fixed $q\left(x\right)$. By the calculus of variations this is equal to

$$\delta L = \frac{dL\left[p(x) + \epsilon\gamma(x), q(x)\right]}{d\epsilon} \Big|_{\epsilon=0} = 0,$$

Therefore we get

$$\begin{aligned} \delta L &= \int \left[\gamma(x)(\ln\left(\frac{p(x)}{q(x)}\right) + 1) + \lambda\delta R(p(x))\gamma(x)\right] dx \\ &= \delta C(p(x))\gamma(x) + \lambda\delta R(p(x))\gamma(x). \end{aligned} \tag{32}$$

Note that, for a fixed $q(x)$, we write the Taylor series expansion as:

$$L(p+\epsilon) = L(p) + \delta L(p)\epsilon + O(\epsilon^2). \tag{33}$$

Substituting equation (30) for a fixed $q(x)$ and equation (32) into equation (33) we get

$$L(p+\epsilon) = C(p) + \lambda R(p) + \delta C(p)\epsilon + \lambda\delta R(p)\epsilon + O(\epsilon^2). \tag{34}$$

However, for a solution of the problem, i.e. a minimizer p^* of $L(p)$

$$dL(p^*) = \delta C(p^*) + \lambda\delta R(p^*) = 0.$$

As a result, changes (errors) in computing the minimum should affect computations up to second order; i.e., $O(\epsilon^2)$.

Notice that this is not a particular characteristic of CIMDO but common to any other estimation method based on optimization techniques, i.e. least squares.

4.2 Robustness w.r.t. changes in the prior distribution

When CIMDO-recovered distribution variations are computed with respect to perturbations in the prior density, we claim that CIMDO-recovered distributions are more robust. We define robustness in the sense that CIMDO-recovered density variations to perturbations in the prior density are smaller than the variations of other standard methods to perturbations in the prior density.

In order to formalize these ideas, recall that CIMDOs functional is expressed in equation (31) as

$$L\left[p,q\right]=\int\left\{p(x)\left[\ln p\left(x\right)-\ln q\left(x\right)\right]+\lambda R\left[p\left(x\right)\right]\right\}dx.$$

In order to compute the variations of the CIMDO-recovered distribution with respect to perturbations in the prior density, we compute the first variation of equation (31) with respect to $q\left(x\right)$ for a fixed $p\left(x\right)$. By the calculus of variations this is equal to

$$\frac{\delta L\left[p,q\right]}{\delta q}=\frac{dL\left[p(x),q(x)+\epsilon\gamma(x)\right]}{d\epsilon}\Big|_{\epsilon=0}=0,$$

therefore, we get,[32]

$$\begin{aligned}\frac{\delta L\left[p,q\right]}{\delta q} &= \frac{d}{d\epsilon}\int -p(x)\ln(q(x)+\epsilon\gamma(x))dx \\ &= \int-\left(\frac{p(x)}{q(x)}\right)\gamma(x)dx.\end{aligned} \tag{35}$$

In order to compare the previous result with other optimization frameworks, consider, instead of the entropy formulation, the following measure of deviation from the prior distribution[33]

$$E_r\left[p,q\right]=\int\left|p(x)-q(x)\right|^r dx. \tag{36}$$

Assuming that restrictions of the form expressed in equation (29) are imposed on this measure of deviation, we get the functional

$$L_r(p,q)=\int\left[\left|p(x)-q(x)\right|^r+\lambda R(p(x))\right]dx. \tag{37}$$

In order to compute the variations of the alternative optimization frameworks with respect to perturbations in the prior density, we compute the first variation of equation (37) with respect to $q\left(x\right)$, for a fixed $p\left(x\right)$. By the calculus of variations, this is equal to

$$\frac{\delta L_r\left[p,q\right]}{\delta q}=\frac{dL_r\left[p(x),q(x)+\epsilon\gamma(x)\right]}{d\epsilon}\Big|_{\epsilon=0}=0,$$

therefore, we get

$$\begin{aligned}\left|\frac{\delta L_r\left[p,q\right]}{\delta q}\right| &= \left|\frac{dL_r}{d\epsilon}\left[p(x),q(x)+\epsilon\gamma(x)\right]\Big|_{\epsilon=0}\right| \\ &\leq \int\gamma\left[r\left|p(x)-q(x)\right|^{r-1}\right]dx.\end{aligned} \tag{38}$$

Note that in this formulation we are considering the absolute value since we are only interested in comparing the magnitude of the deviations.

[32] Since the restriction does not depend on q, it plays no role when computing this variation using the Euler-Lagrange equation, so in this case $\frac{d}{dq(x)}C(p(x),q(x))=\frac{d}{dq(x)}L(p(x),q(x))$.

[33] In this measure of deviation, we assume that $r\geqslant 2$ is a natural choice, that is consistent with common utility (loss) functions. $r<1$ would lead to unnatural situations. By unnatural we mean that $r<1$ is not a well defined norm; e.g., the triangle innequality is not valid anymore. However, for values $1\leq r<2$ a direct comparison is not possible and more detailed analysis of the constants appearing in the estimates would be required. Therefore in the range $1\leq r<2$ the proof presented is not conclusive.

In order to demonstrate that CIMDO is more robust than alternative optimization frameworks with respect to perturbations in the prior distribution, we prove that the result presented in equation (38) is greater than or equal to the result presented in equation (35).

Proof. $\frac{\delta L_r[p,q]}{\delta q} \geq \frac{\delta L[p,q]}{\delta q}$.
Equation (35) can be expressed as

$$E = \frac{\delta L\,[p,q]}{\delta q} = \int \left| \left(\frac{p(x)}{q(x)} \right) \gamma(x)^{1/2} \right| \gamma(x)^{1/2} dx. \tag{39}$$

By Hölder's inequality, we get,[34]

$$E \leqslant \left[\int \gamma(x)^{\frac{r-1}{2}} \left| \frac{p(x)}{q(x)} \right|^{r-1} dx \right]^{\frac{1}{r-1}} \left[\int \gamma(x)^{\frac{r-1}{2r}} dx \right]^{\frac{r}{r-1}}. \tag{40}$$

Recall that it is assumed that is $\gamma(x)$ is bounded and has compact support; therefore, all the integrals are taken effectively on finite intervals and not in $(-\infty, \infty)$. As a result, the second integration on the right hand side of inequality (40) is $\leqslant C_0$.

$$E \leqslant C_0 \left[\int \gamma(x)^{\frac{r-1}{2}} \left| \frac{p(x)}{q(x)} \right|^{r-1} dx \right]^{\frac{1}{r-1}}.$$

Due to the fact that $\gamma(x)$ is bounded and has compact support, we get for some constant $\widetilde{C}$, that

$$\begin{aligned} E &\leqslant \widetilde{C} \left[\int \left| \frac{p(x)}{q(x)} \right|^{r-1} dx \right]^{\frac{1}{r-1}} \\ &= \widetilde{C} \left[\int \left| \frac{p(x)}{q(x)} - 1 + 1 \right|^{r-1} dx \right]^{\frac{1}{r-1}}. \end{aligned}$$

By the triangle inequity, we get,[35]

$$E \leqslant \widetilde{C} \left[\int \left[\left| \frac{p(x)}{q(x)} - 1 \right| + 1 \right]^{r-1} dx \right]^{\frac{1}{r-1}}.$$

Finally, by Minkowsky's inequality, we get,[36]

$$E \leqslant \widetilde{C} \left[\int \left[\left| \frac{p(x)}{q(x)} - 1 \right| \right]^{r-1} dx \right]^{\frac{1}{r-1}} + \int \left[1^{r-1} \right] dx^{\frac{1}{r-1}},$$

Recall that all the integrals are effectively performed in a compact set. Therefore the second integral on the right hand side is equal to $\check{C}$, thus we get,

$$= \widetilde{C} \left[\int \left[\left| \frac{p(x)}{q(x)} - 1 \right| \right]^{r-1} dx \right]^{\frac{1}{r-1}} + \check{C}. \tag{41}$$

Equivalently, equation (38) can be expressed as

[34] Hölder's inequality establishes that $\int ab \leqslant \left(\int a^m\right)^{1/m} \left(\int b^n\right)^{1/n}$ where $\frac{1}{m} + \frac{1}{n} = 1$. If we set $m = r-1$, then $\frac{1}{m} = \frac{1}{r-1}$ and $\frac{1}{n} = 1 - \frac{1}{r-1} \Longrightarrow n = \frac{r-1}{r}$

[35] The triangle inequality establishes $|a+b| \leqslant |a| + |b|$ and $|a-b| \geqslant |a| - |b|$.

[36] Minkowsky's inequality establishes that $\left[\int (a+b)^n\right]^{1/n} \leq \left[\int a^n\right]^{1/n} + \left[\int b^n\right]^{1/n}$.

$$\frac{\delta L_r [p,q]}{\delta q} = \int \left[r \left| q(x) \right|^{r-1} \left| \frac{p(x)}{q(x)} - 1 \right|^{r-1} \gamma(x) \right] dx. \tag{42}$$

By the triangle inequality, we get,

$$\geqslant \int r \left| q(x) \right|^{r-1} \left[\left| \frac{p(x)}{q(x)} \right| - 1 \right]^{r-1} \gamma(x) dx,$$

Since $\left| q(\bar{x}) \right|^{r-1} \gamma(\bar{x})$ is a positive value, by the mean value theorem for integrals, where $\bar{x}$ is the value guaranteed by this theorem, see Courant and John (1965), we get,

$$= \left| q(\bar{x}) \right|^{r-1} \gamma(\bar{x}) \int r \left[\left| \frac{p(x)}{q(x)} \right| - 1 \right]^{r-1} dx.$$

$$\geqslant C \int r \left[\left| \frac{p(x)}{q(x)} \right| - 1 \right]^{r-1} dx, \tag{43}$$

where $0 < C \leq \left| q(\bar{x}) \right|^{r-1} \gamma(\bar{x})$.

From these results, we can see that the expression in equation (42) is greater than or equal to the expression in equation (43). We can also see that the expression in equation (43) is greater than the expression in equation (41) and that the expression in equation (41) is greater than the expression in equation (39). Therefore we get

$$\begin{aligned} \frac{\delta L_r [p,q]}{\delta q} &\geqslant C \int r \left[\left| \frac{p(x)}{q(x)} \right| - 1 \right]^{r-1} dx \\ &\geqslant \widetilde{C} \left[\int \left[\left| \frac{p(x)}{q(x)} - 1 \right| \right]^{r-1} dx \right]^{\frac{1}{r-1}} + \check{C} \\ &\geqslant \frac{\delta L [p,q]}{\delta q} \qquad \blacksquare \end{aligned}$$

Recall that this proof is valid when $r \geqslant 2$. Note that this difference becomes larger with r. [37]

5 CIMDO: empirical application

In this section, we present an empirical exercise that shows how the CIMDO methodology is implemented in order to recover a portfolio distribution containing two types of loans held by unlisted firms with different ratings.

Using the same information set, we calibrate the multivariate normal distribution, multivariate t-distribution and a mixture of normals, which are the most common parametric distributions assumed under the structural approach to model loans credit risk.

Although the calibration procedures that we apply in this section are in line with common practice by risk modelers, we would like to emphasize that we believe that such procedures are not adequate. This is because we consider that the parametric distributions used under the structural approach cannot possibly be calibrated correctly under the data limitations binding the portfolio credit risk modelling of SMEs and unlisted firms.

This claim is proved at the end of this section when under the PIT criterion the CIMDO-recovered density outperforms the competing parametric distributions.

[37] Note that the term $\frac{1}{r-1}$ is by definition < 1 and that the constant $\check{C}$ does not depend on r.

5.1 Information set

In our dataset, loans given to SMEs and unlisted firms are classified and aggregated according to their credit-risk rating. For a given risk-rating, the empirical probabilities of default are recorded at each period of time. Thus, the probabilities of default reflect the percentage of borrowers, classified under a given risk-rating, that defaulted at a given point in time.

This database was provided by the Comision Nacional Bancaria y de Valores (CNBV), the Mexican Financial Regulatory Agency. Banks operating in Mexico, on the basis of a rating system set out by the regulatory authority, determine ratings internally and then report them to the CNBV.[38]

In order to infer the CIMDO-recovered distribution, we use the annual probabilities of default of two types of Mexican unlisted firms, classified with different credit-risk ratings. These ratings belong to two different "below-investment grade" classes, according to the rating system defined by the CNBV. Our dataset covers the period 1993-2000, and we use the PoDs corresponding to 1996, the peak of the Mexican financial crisis.

Note that, since classification and aggregation of loans can be carried out according to other loan characteristics, such as sectoral activity of borrowers, geographical location of borrowers, type of collateral backing up the loan etc., the proposed procedure can easily be implemented using the PoDs of loans aggregated under any of these classification schemes.

With this database, it was possible to compute the through-time-average probability of default and, as a result, the through-time-average default-threshold as indicated in equation (16) for each type of loan. The probabilities of default and the default-thresholds that were used are presented in Table 1.

Table 1: Information set

Variable\Asset Type	**Loan x**	**Loan y**
Through-time-average probability of default ($\overline{PoD^m}$)	.15	.19
Default-threshold (X_d^m)	1.0364	.8779
Annual Probability of Default 1996 (PoD_{96}^m)	.22	.29

Note that this is the information set that is usually available for modelling loan portfolio multivariate distributions for bank regulators and other modelers working for institutions with *arm-length relationships* with firms.

5.2 CIMDO: density recovery

In order to obtain the values of the Lagrange multipliers that define the CIMDO-recovered multivariate density, we solve the system of equations composed by the restrictions imposed on the constrained optimization problem presented in equation (20). This system of equations was set as follows

$$\begin{aligned} \int_{-\infty}^{\infty}\int_{1.0364}^{\infty}\widehat{p(x,y)}dxdy &= .22, \\ \int_{-\infty}^{\infty}\int_{.8779}^{\infty}\widehat{p(x,y)}dydx &= .29, \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\widehat{p(x,y)}dxdy &= 1, \end{aligned} \tag{44}$$

where we make use of the default thresholds X_d^x, X_d^y and the empirical probabilities of default PoD_{96}^x, PoD_{96}^y for each type of borrower, and where $\widehat{p(x,y)}$ is defined as in equation (22).

Solving numerically for the Lagrange multipliers $\widehat{\mu}$, $\widehat{\lambda}_1$, $\widehat{\lambda}_2$, we recover CIMDOs distribution. The solution to the Lagrange multipliers is presented in Table 2.

[38] The Mexican rating methodology is described in: http://www.cnbv.gob.mx

Table 2: Lagrange multipliers

Lagrange multipliers	**Value**
$\widehat{\mu}$	-.4765
$\widehat{\lambda}_1$	-.5627
$\widehat{\lambda}_2$	-.7783

5.3 Common parametric distributions used in the structural approach

As already explained in Section 2.1, in its most basic version, the structural approach assumes that the firms' logarithmic asset values are normally distributed (Merton,1974). However, the multivariate t-distribution and mixture models have also been widely used.

The t-distribution has been justified by the fact that financial assets' returns exhibit heavier tails than would be predicted by the normal distribution. It is common to observe large events, e.g. extreme asset price movements, more frequently than the normal distribution predicts. Empirical support for modelling univariate returns with t-distributions can be found in Danielsson and de Vries (1997) and Hoskin, Bonti and Siegel (2000), while Glasserman, Heidelberger and Shahabuddin (2000) present the multivariate modelling framework. When it is assumed that the logarithmic asset values of the assets making up a portfolio follow marginal t-distributions with the same degrees of freedom, the modelling of dependence and portfolio credit risk is very similar to the basic version of the structural approach (which assumes a multivariate normal distribution). Still, the modelling of multivariate t-distributions with marginals containing the same degrees of freedom has proved to be too inflexible. Extensions of multivariate t-distributions that allow for different degrees of freedom in their marginals are possible; but, under these assumptions, the multivariate t-distributions are no longer from the elliptical family of distributions.[39] When multivariate distributions abandon the elliptical family of distributions, variance-covariance matrices are no longer sufficient to describe them. In these cases, different paths have been proposed.

One possible path to follow is to infer the dependence structure from historical data; yet, in practice, this is a difficult task, because historical simulation only incorporates information about extreme returns as long as they are included in the sample period. This fact represents a drawback for the use of this methodology, since sufficiently long time horizons of data are necessary to include different stages of the economic cycles (Mina and Xiao, 2001 and Butler and Schachter, 1998).

Mixture models provide an alternative option that has been used for the modelling of loan credit risk. These models assume that a firm's logarithmic asset values are generated from a mixture of two different normal distributions: the distribution of the *quiet* state and the distribution of the *volatile* state. Under the first, logarithmic asset values are assumed to follow a normal distribution with mean μ_1 and variance σ_1, $x \sim N_1(\mu_1, \sigma_1)$ with probability $\mathbf{p_{ro1}}$. Under the latter, logarithmic asset values are assumed to follow a normal distribution with mean μ_2 and variance σ_2, $x \sim N_2(\mu_2, \sigma_2)$ with probability $\mathbf{p_{ro2}}$. Therefore, the probability density function (*pdf*) of the firm's logarithmic asset values is generated as follows:

$$PDF_{Mixture} = \mathbf{p_{ro1}}\left[N_1(\mu_1, \sigma_1)\right] + \mathbf{p_{ro2}}\left[N_2(\mu_2, \sigma_2)\right]. \tag{45}$$

The resulting distributions under these models exhibit heavy tails due to the random nature of the volatility.[40] The calibration of this model is difficult. Note that, in the univariate case, it is necessary to estimate five parameters, two variances, two means and the probability of being in a volatile state. Moreover, the calibration in the multivariate case becomes more difficult, as it is necessary to calibrate two covariance matrices corresponding to the *quiet* and *volatile* states for the multivariate distributions. (McLachlan and Basford, 1987 and Zangari, 1996).

[39] A multivariate distribution belongs to the elliptical family of distributions, if its marginal distributions are assumed to be Normal or t distributions with the same degrees of freedom. Elliptical distributions are distributions whose density is constant on ellipsoids. In two dimensions the contour lines of the density surface are ellipses. Therefore, their variance-covariance matrices describe their dependence structure. See Embrechs, McNeil and Straumann (1999).

[40] These distributions decay at a lower speed than the normal distribution, although they are still exponential.

5.4 Empirical calibration of parametric distributions

Using empirical frequencies of default, we proceed to calibrate the parametric distributions described above. As argued throughout the document, empirical frequencies of default represent only partial information on the marginal distributions of portfolio multivariate distributions, so they do not represent an adequate dataset for proper calibration of multivariate parametric distributions. However, given that PoDs are the only statistic available for SMEs and unlisted companies, risk modelers usually have no other choice than to calibrate parametric distributions in a way that ensures that empirical probabilities of default "fit" into the default region of the marginal distributions of the chosen multivariate distribution.

The most common assumption is that firms' logarithmic asset values conditioned on the period's level of volatility are independent across time and follow a normal distribution or a t-distribution (Mina and Xiao, 2001). Modelers in the industry usually calibrated these distributions by adjusting the volatility parameter in order to ensure that the region of default of the assumed parametric distributions equals the empirically observed probabilities of default.[41] Following this procedure, we found the volatility values by fitting the following equations for each type of loan.

$$\begin{aligned} \int f(x|\sigma_t)\chi_{[X_d^x,\infty)}dx &= PoD_{96}^x, \\ \int f(y|\sigma_t)\chi_{[X_d^y,\infty)}dy &= PoD_{96}^y, \end{aligned} \tag{46}$$

where $f(x|\sigma_t)$ represents the conditional normal or the non-standard t-distribution, calibrated with their respective volatility (standard deviation) parameters σ_t. We assumed a t-distribution with degrees of freedom v=6. Empirical evidence presented in Hansen (1994) indicates that this is a reasonable assumption.

Volatility values are presented in Table 3.

Table 3: Volatility values

Distribution\Asset Type	**Asset x**	**Asset y**
Normal Distribution σ_t	1.3422	1.5864
$t_{v=6}$ -distribution σ_t	1.5353	1.8386

The univariate restrictions, defined in equation (46), can be expressed as restrictions imposed on the multivariate densities' marginals. Mathematically, we formalize this as

$$\begin{aligned} \int\int f(x,y|\sigma_t)\chi_{[X_d^x,\infty)}dxdy &= PoD_{96}^x, \\ \int\int f(x,y|\sigma_t)\chi_{[X_d^y,\infty)}dydx &= PoD_{96}^y, \end{aligned} \tag{47}$$

where $f(x,y|\sigma_t)$ represents the bivariate conditional normal distribution or the bivariate non-standard t-distribution. Covariances are set equal to zero because in the case of SMEs and non-listed firms, information constraints usually prevent modelers from defining a dependence structure.

In the case of the mixture model, the assumed probability density function is defined in equation (45). In order to calibrate this distribution in the bivariate case, we fitted the following equation

$$Pdf_{Mixture} = \int_{X_d^y}^{\infty}\int_{X_d^x}^{\infty}\left\{\mathbf{p_{ro1}}\left[N_1(\mu_1,\Sigma_1)\right] + \mathbf{p_{ro2}}\left[N_2(\mu_2,\Sigma_2)\right]\right\}dxdy = \mathbf{PoD}_{96}.$$

Where $\mathbf{PoD}_{96}^M$ is a vector containing the frequency of default for the borrowers in the portfolio, $\mathbf{p_{ro1}}$, $\mathbf{p_{ro2}}$ are the values indicating the probabilities of the *quiet* and *volatile* states, N_1, N_2 are bivariate normal

[41] Due to the lack of data to compute dependence structures for SME's and non-quoted firms, most of the times, it is assumed that the random variables in the calibrated multivariate distributions are independent. This has been observed in my experience as a banking regulator and a modeller in the private sector.

distributions under the *quiet* and *volatile* states, μ_1, μ_2 are the mean borrowers' asset values under the *quiet* and *volatile* states and Σ_1, Σ_2 are variance covariance matrices for the bivariate distribution under the *quiet* and *volatile* states. The values of these parameters are presented in Table 4.

Table 4: Mixture model parameters

Parameter\State	**Quiet**	**Volatile**
Probability	$\mathbf{p_{ro1}} = .7817$	$\mathbf{p_{ro2}} = .2183$
μ_1 μ_2	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} .3000 \\ .3000 \end{bmatrix}$
Σ	$\begin{bmatrix} 1.0000 & 0 \\ 0 & 1.5104 \end{bmatrix}$	$\begin{bmatrix} 100.0000 & 0 \\ 0 & 109.1398 \end{bmatrix}$

When fitting the parameters of this equation we tried to get values as close as possible to the assumptions that are usually made for the specification of this type of model (McLachlan and Basford, 1987). Furthermore, when calibrating these parameters, we remain consistent with the fact that asset y is riskier than asset x under both the *quiet* and *volatile* states.

5.5 Density evaluation under the PIT

As we discussed above, for measurement of loan credit risk, common practice involves the calibration of parametric distributions in a way that ensures that empirical probabilities of default "fit" into the default region of the distribution in use. We show that, under the PIT criterion, the CIMDO-recovered distributions improve over these parametric distributions. We claim that this fact provides evidence to assert that CIMDO reduces the distribution's specification risks, when data limitations binding the portfolio credit risk measurement of loans granted to SMEs and unlisted firms do not allow for a proper calibration of the parametric distributions that are usually assumed in portfolio credit risk.

As we proved in Section 3, the series of Probability Integral Transformations of a series of density forecasts are distributed **iid** $\mathbf{U}(\mathbf{0}, \mathbf{1})$ if the density forecast coincides with the true DGP. *Thus, to assess whether a series of density forecasts coincides with the corresponding series of true conditional densities, it is only necessary to test whether the sequence of density forecast PITs are* **iid** $\mathbf{U}(\mathbf{0}, \mathbf{1})$.

However, given the highly restrictive data sets for SMEs and unlisted firms, one would not expect to obtain a series of PITs that are **iid** $\mathbf{U}(\mathbf{0}, \mathbf{1})$. We therefore *focused on the relative improvements of one specification versus another* by analyzing how close their series of probability integral transformations are to **iid** $\mathbf{U}(\mathbf{0}, \mathbf{1})$. We followed this procedure to compare CIMDO-recovered densities versus the competing parametric densities that were calibrated in the previous section, i.e. the conditional normal distribution, the conditional t-distribution and the mixture model.

We also compared CIMDO with the standard normal distribution. This comparison was carried out in order to envisage the degree of misspecification that could be reached when the probability of default under the prior is completely inconsistent with empirical probabilities of default at each period of time. This could be the case if information restrictions completely preclude any form of calibration, even the imperfect calibration procedures that are undertaken currently. Interestingly, the same type of misspecification could be reached if it is assumed that *through-time* probabilities of default are the statistics that should be used for the computation of portfolio multivariate densities that are used to calculate the PLD distribution of a loan portfolio.

To perform the density evaluation, under the PIT criterion, we proceeded as follows.

1. **Simulation of data generating process (DGP)**

Given that for SMEs and unlisted firms we just have information on the probabilities of default and not on the actual realizations of their underlying asset values, it was necessary to assume the data generating that was driving the firms' asset prices. There is an infinite number of specifications that could have been chosen;[42] nonetheless, we chose a data generating process that was consistent with the empirically observed probabilities of default but also which mimic, in the best possible way, the behavior of asset returns in emerging markets in periods of economic distress (recall that we are using information related to 1996, which was a time of financial distress in Mexico). The specification that we chose was

$$\mathbf{f}(\mathbf{x}, \mathbf{y}) \sim t_{v=6} \left\{ \begin{bmatrix} .3613 \\ .4004 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right\}, \tag{48}$$

where $\mathbf{f}(\mathbf{x}, \mathbf{y})$ is the bivariate t-distribution with $v = 6$ and non-central parameter $\omega^x = .3613$ and $\omega^y = .4004$.[43]

The justification of the appropriateness of this distributional assumption for financial data is presented in Hansen (1994). Of course, for the specific case of SMEs and unlisted firms in reality we have no information which would enable us even to attempt to mimic in an adequate manner the behavior of the asset returns of these types of firms. Nonetheless, this specification seems reasonable,[44] and is consistent with empirical evidence provided by Bekaert et al (2002), who analyze the empirical behavior of the prices of different types of financial assets in periods of financial distress in emerging markets.[45]

Under this specification, we simulated a bivariate series of 10,000 observations.

2. **Decomposition of multivariate distribution**

Once the DGP was simulated, we proceeded to decompose the bivariate distribution in two different ways, as we stated in equations (24) and (25). Note that each case can be factorized into the product of its conditional and marginal distributions, therefore we end up with 1 series of conditional and 1 series of marginal distributions (2 series) for each case.

3. **Computation of probability integral transformations (PITs)**

We then computed the probability integral transformation (PIT's) of the random variable realizations under the following conditional and marginal distributions,

$$z_{x|y} = P(x|y), \qquad z_y = P(y), \qquad z_{y|x} = P(y|x), \qquad z_x = P(x).$$

where P represents the *cdf* of each of the evaluated distributions.

We henceforth refer to the PITs of the random variable realizations under the conditional/marginal densities as: *z-variables.*

We obtained the *z-variables* for each of the distributional specifications that we evaluated and compared; these being the standard normal, the conditional normal, the conditional t-distribution, the mixture model and CIMDO. We henceforth refer to the PITs obtained under each of these parametric distributions as: NStd, NCon, TCon, NMix and CIMDO respectively.

4. **Testing for iid $\mathbf{U(0,1)}$**

Following Diebold et al (1999), our aim is to assess whether a bivariate density specification is correct by testing whether the series $z_{x/y}$, z_y, $z_{y/x}$, z_x is **iid $\mathbf{U(0,1)}$** individually and also when taken as a whole. Testing for **iid $\mathbf{U(0,1)}$** is usually done in two stages. First the *iid* assumption is tested and then, conditional on the series being *iid*, the uniform distribution is tested.

[42] Recall that in contrast to the simulation excercise presented in Segoviano 2005a, where our interest focused on testing the entropy-derived density under alternative assumptions of the data generating processes driving the firms' asset values, in this chapter, our interest focused on comparing the entropy-derived density with the distributions that are most commonly used for portfolio credit risk modelling. Thus, in order to make the excercise manageable, we fixed the data generating process that was assumed to drive the firms' asset values. We tried to choose the assumption that seemed the most appropriate.

[43] The univariate t-distribution is defined as $t(v) = \frac{z}{\sqrt{x/v}}$ where z is a $N(0,1)$ variable and x is a $\chi^2(v)$ independent of z. If the mean of the normal distribution is not 0, then the ratio has the noncentral t-distribution with non-central parameter ω. The most general representation of the noncentral t-distribution is quite complicated. Johnson and Kotz (1970) give a formula for this distribution.

[44] Because, as previously noted, the t-distribution is justified due to the fact that it has been observed that the distributions of financial assets' prices show heavier tails than would be predicted by the normal distribution.

[45] In additional Monte Carlo experiments that we performed, the t-distribution with $\nu = 6$ represented the "toughest assumption" to be outperformed by the CIMDO-recovered distributions. This was an additional reason why we were interested in performing density evaluation under this assumption.

Test for the iid assumption

As we proved in Section (3), the PITs of the marginal and conditional densities in which a multivariate distribution can be decomposed, i.e. the series $z_{x/y}$, z_y, $z_{y/x}$, z_x, are independent *at a given period of time*; therefore, there is no need to test for independence.[46]

Test for the uniform distribution assumption

Conditional on the series being independent, we proceeded to test for the uniform distribution assumption.

Note that if $F(z) \sim \mathbf{U(0,1)}$ and $z \in [0,1]$, then $F(z) = z$ for $z \in [0,1]$. Thus the *cdf* of z is a 45° degree line. Equivalently, the *cdf* of the PITs of the random variable realizations under the true *DGP* will be a 45° degree line if they are uniformly distributed. *Thus, in order to assess the correctness of a given density specification, we plotted the empirical cdf's of their z-variables and checked how closely these were to the 45° degree line.* Then we proceeded to compute Kolmogorov-Smirnov (K-S) tests with H_0: $F = \mathbf{U(0,1)}$, H_a: $F \neq \mathbf{U(0,1)}$ at the 5% significance level. K-S tests results are presented in Appendix 4.

5. **Results:**

Results showing the empirical *cdf's* of the *z-variables* are presented in Figures 3 to 14.

[46] In time series frameworks, auto-correlation coefficients (ACFs) and Ljung-Box-Q-statistics of the *z-variables* are used to test for independence. At a given period of time, the *z-variables* are independent, as was proved in Section (3). Therefore, it is not necessary to test for independence.

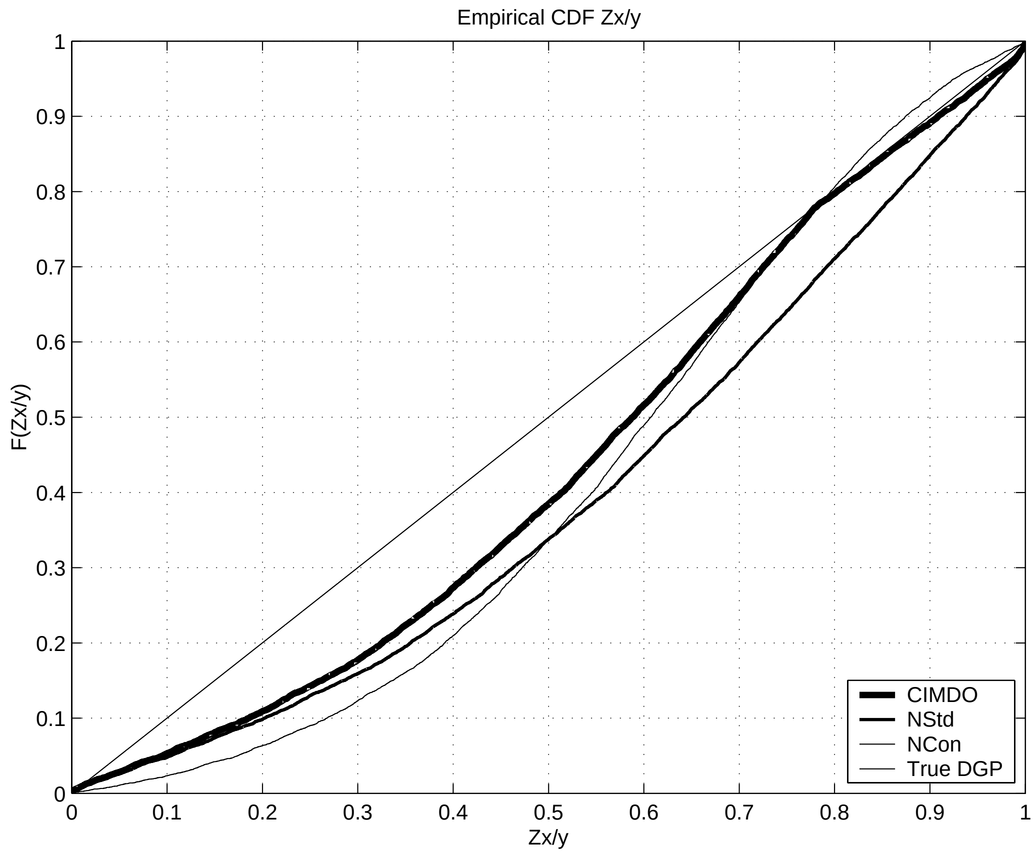

Figure 3: Empirical CDF Zx|y: CIMDO, NStd, NCon

In this figure, we compare the empirical cdf's of the $z_{x|y}$ variables corresponding to CIMDO, NStd, NCon and the true-DGP (45° degree line). It is clear that CIMDOs distribution outperforms NStd and NCon along the whole integration domain of the distribution. However, it is in the region of default (upper right corner) where improvements are the best.

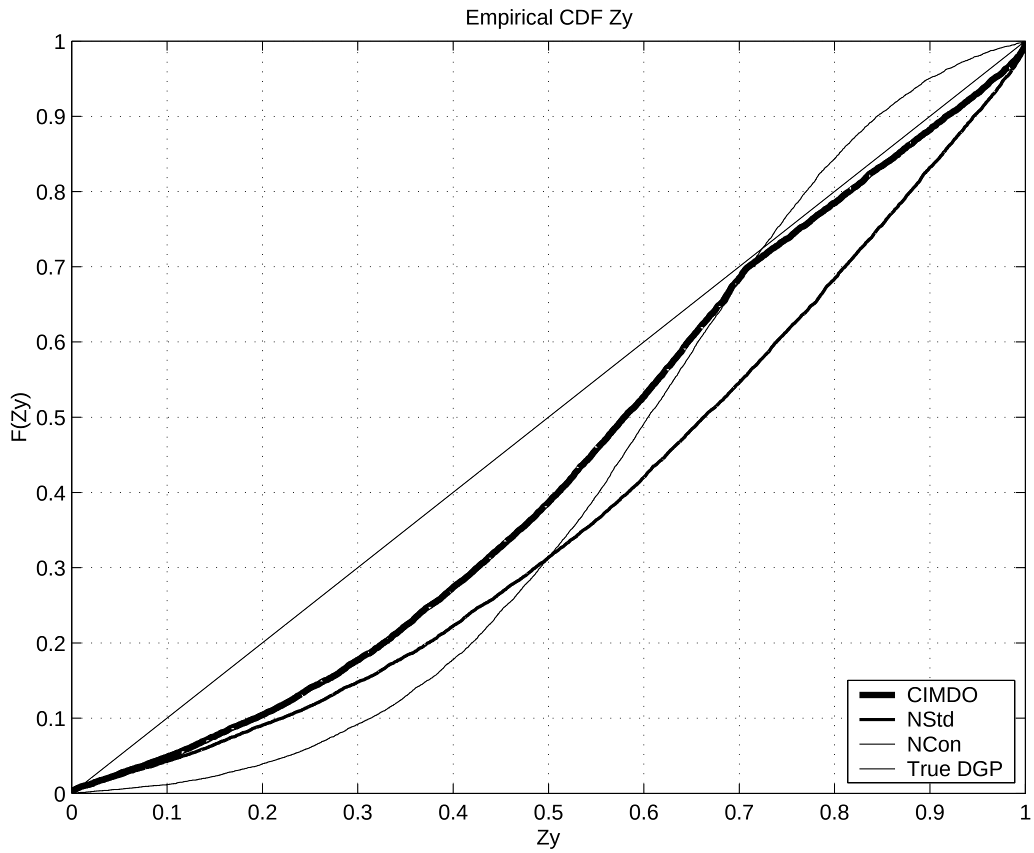

Figure 4: Empirical CDF Zy: CIMDO, NStd, NCon

In this figure, we compare the empirical cdf's of the z_y variables corresponding to CIMDO, NStd, NCon and the true-DGP (45° degree line). It is clear that CIMDOs distribution outperforms NStd and NCon along the whole integration domain of the distribution. However, it is in the region of default (upper right corner) where improvements are the best.

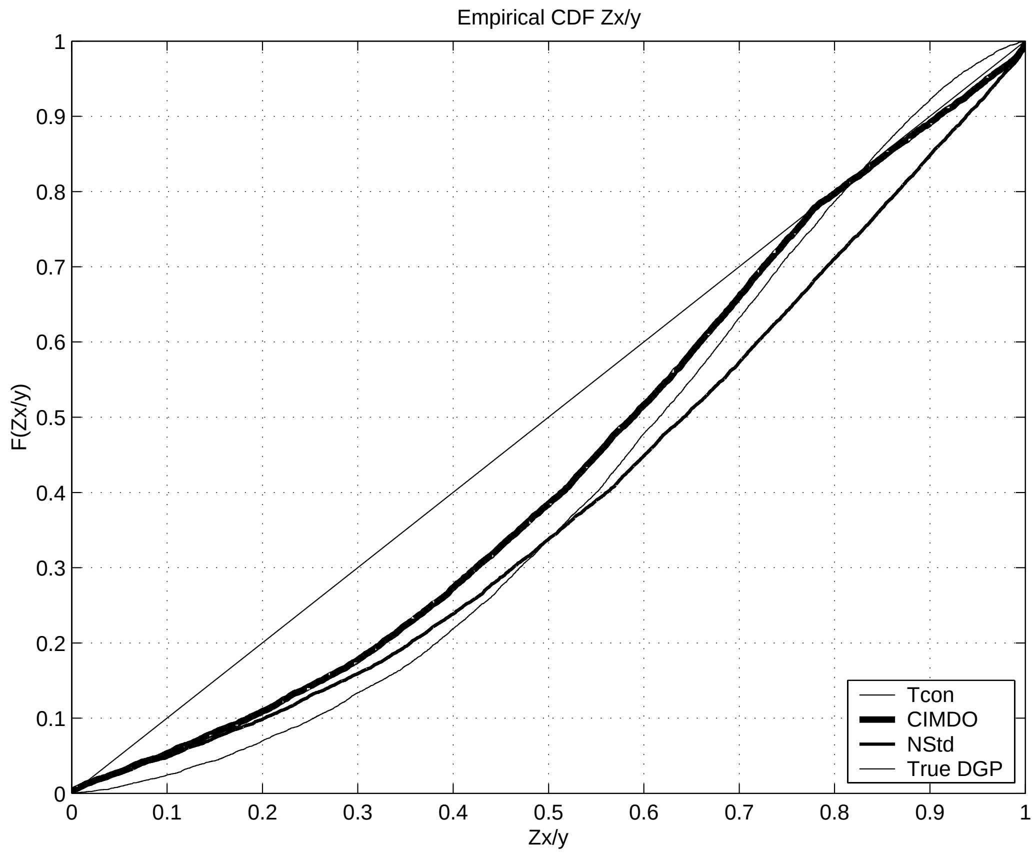

Figure 5: Empirical CDF Zx|y: CIMDO, NStd, TCon

In this figure, we compare the empirical cdf's of the $z_{x|y}$ variables corresponding to CIMDO, NStd, TCon and the true-DGP (45° degree line). It is clear that CIMDOs distribution outperforms NStd and TCon along the whole integration domain of the distribution. However, it is in the region of default (upper right corner) where improvements are the best.

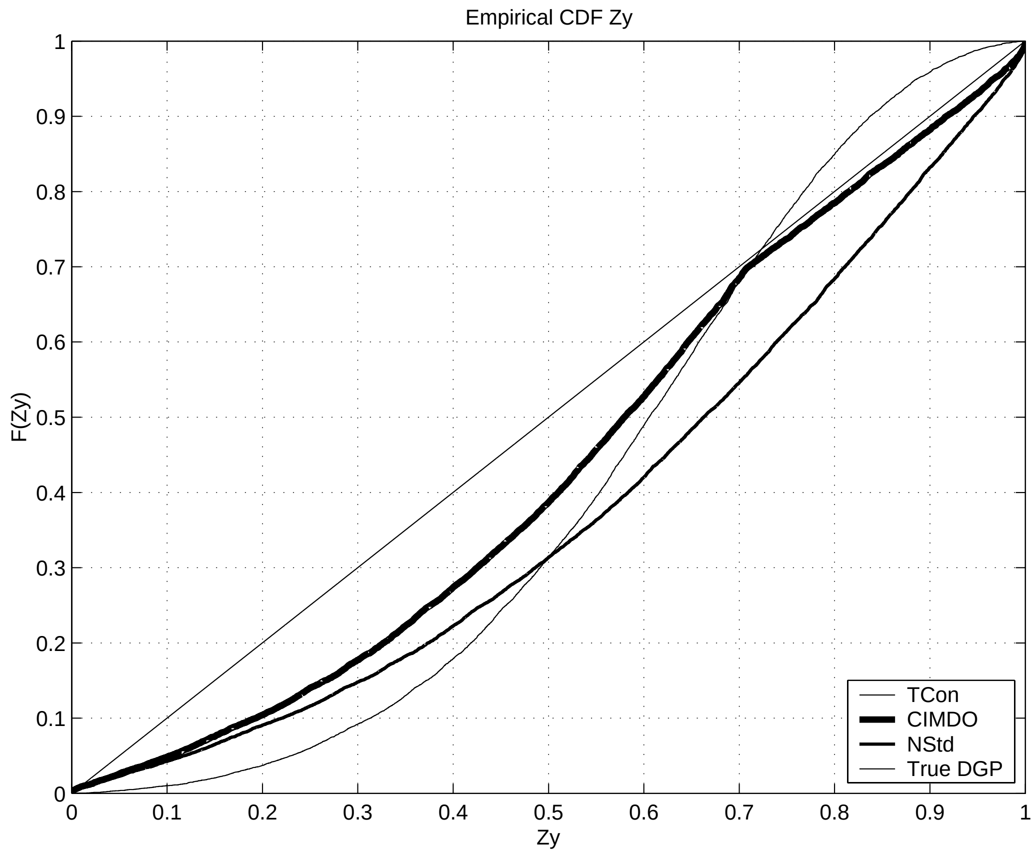

Figure 6: Empirical CDF Zy: CIMDO, NStd, TCon

In this figure, we compare the empirical cdf's of the z_y variables corresponding to CIMDO, NStd, TCon and the true-DGP (45° degree line). It is clear that CIMDOs distribution outperforms NStd and TCon along the whole integration domain of the distribution. However, it is in the region of default (upper right corner) where improvements are the best.

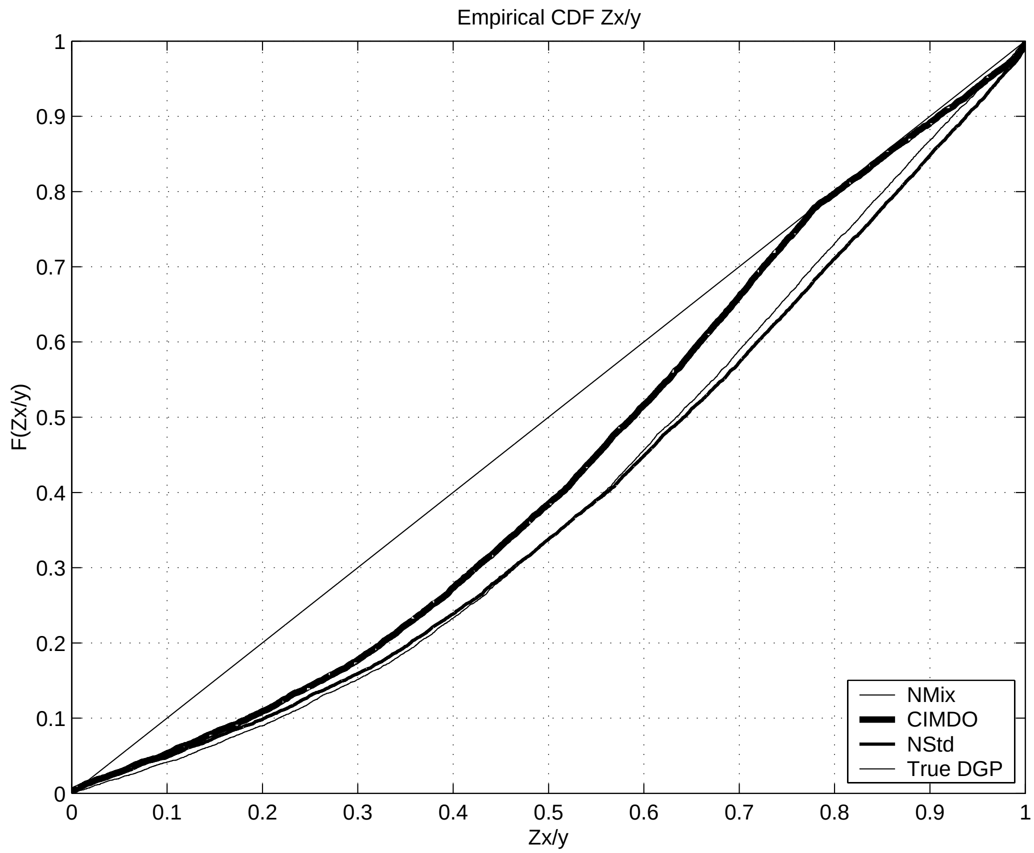

Figure 7: Empirical CDF Zx|y: CIMDO, NStd, NMix

In this figure, we compare the empirical cdf's of the $z_{x|y}$ variables corresponding to CIMDO, NStd, NMix and the true-DGP (45° degree line). It is clear that CIMDOs distribution outperforms NStd and NMix along the whole integration domain of the distribution. However, it is in the region of default (upper right corner) where improvements are the best.

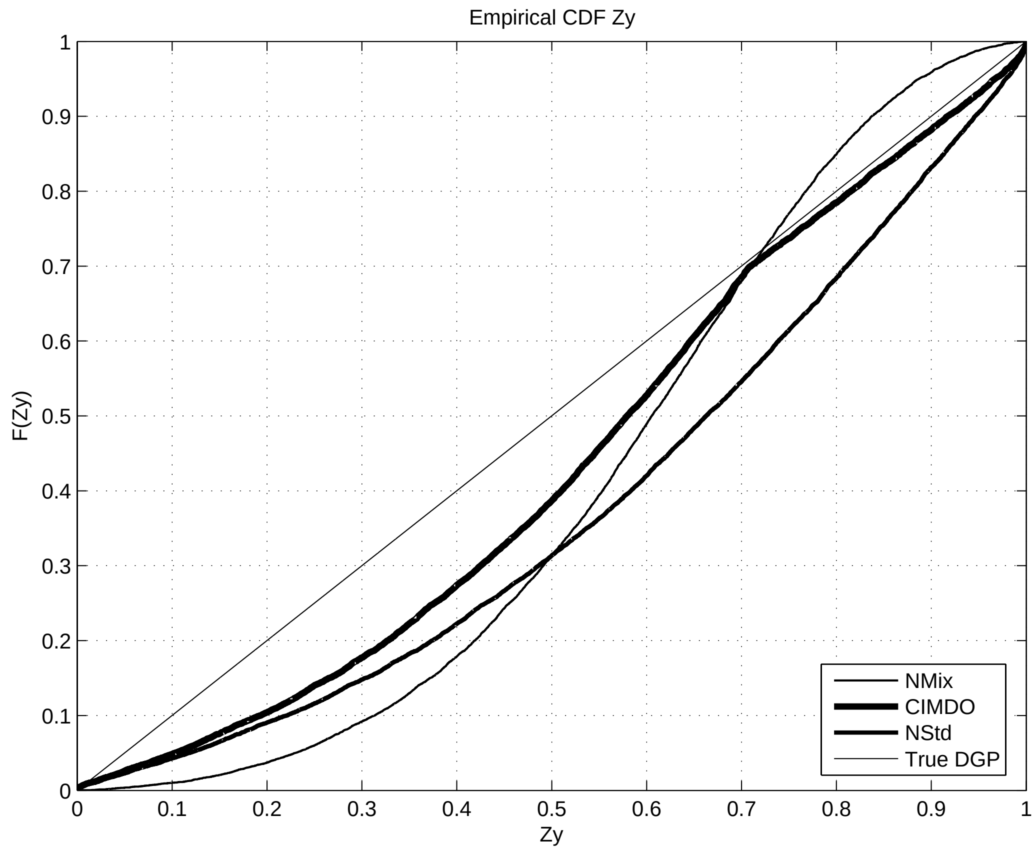

Figure 8: Empirical CDF Zy: CIMDO, NStd, NMix

In this figure, we compare the empirical cdf's of the z_y variables corresponding to CIMDO, NStd, NMix and the true-DGP ($45°$ degree line). It is clear that CIMDOs distribution outperforms NStd and NMix along the whole integration domain of the distribution. However, it is in the region of default (upper right corner) where improvements are the best.

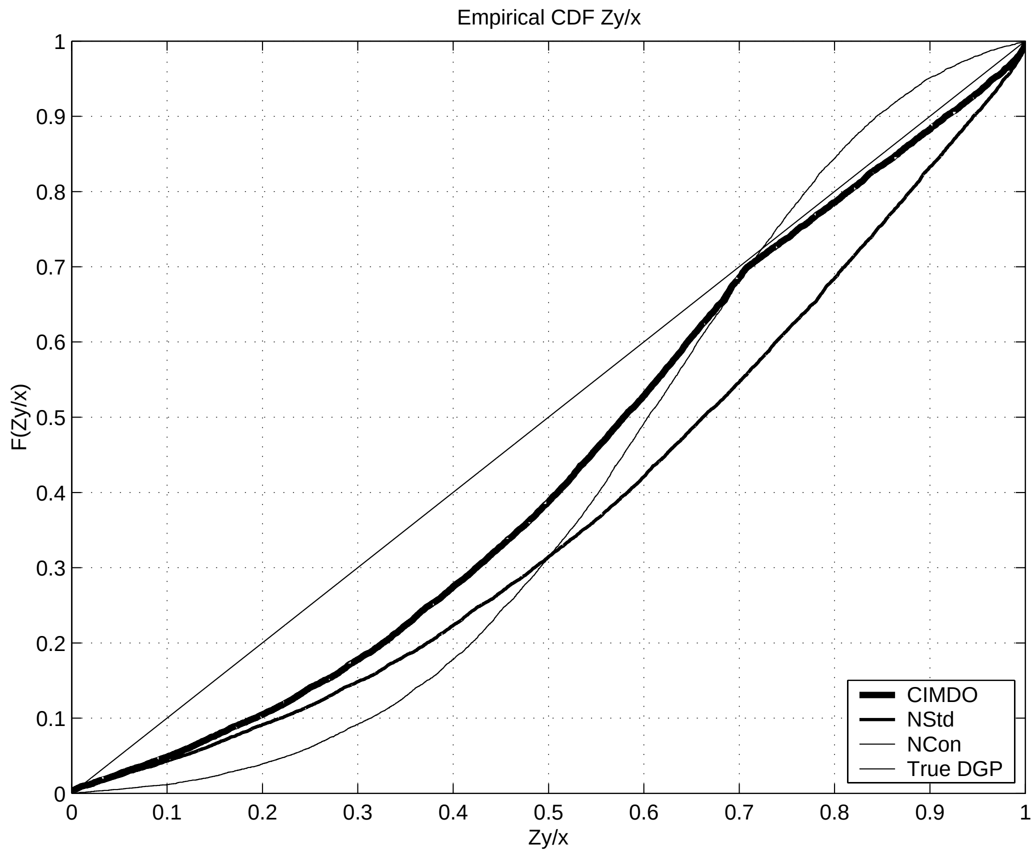

Figure 9: Empirical CDF Zy|x: CIMDO, NStd, NCon

In this figure, we compare the empirical cdf's of the $z_{y|x}$ variables corresponding to CIMDO, NStd, NCon and the true-DGP (45° degree line). It is clear that CIMDOs distribution outperforms NStd and NCon along the whole domain of the distribution. However, it is in the region of default (upper right corner) that the improvements are the best.

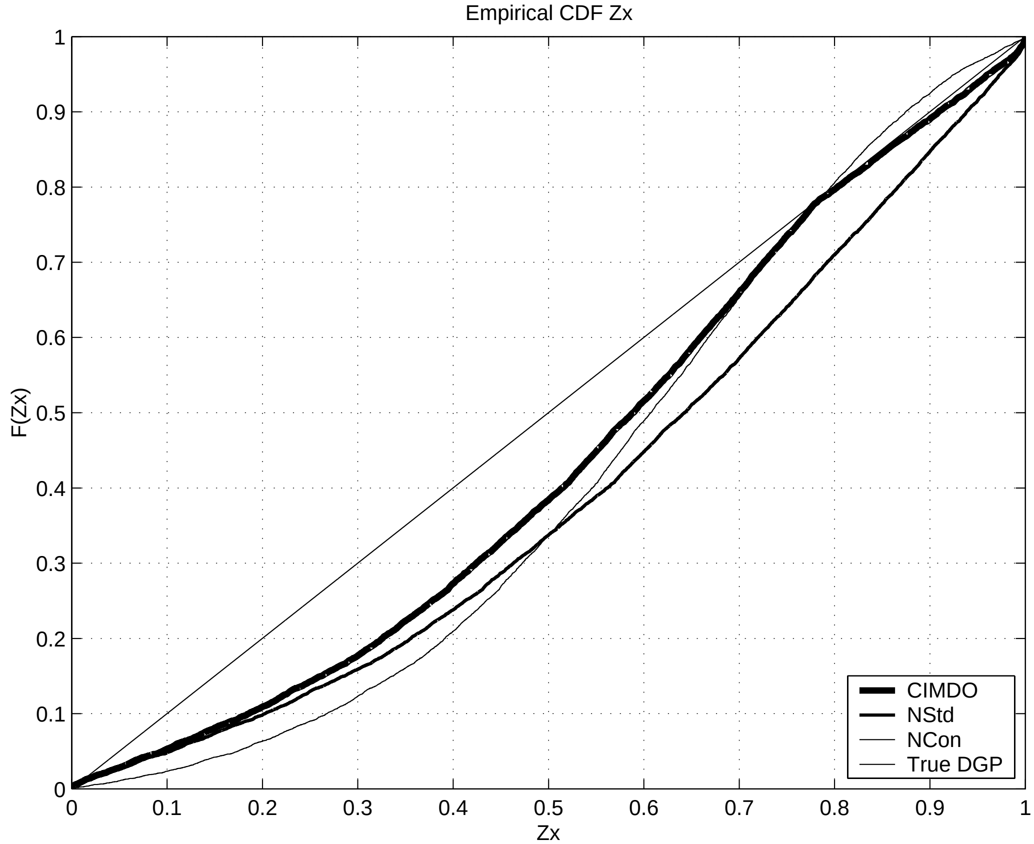

Figure 10: Empirical CDF Zx: CIMDO, NStd, NCon

In this figure, we compare the empirical cdf's of the z_x variables corresponding to CIMDO, NStd, NCon and the true-DGP ($45°$ degree line). It is clear that CIMDOs distribution outperforms NStd and NCon along the whole domain of the distribution. However, it is in the region of default (upper right corner) that the improvements are the best.

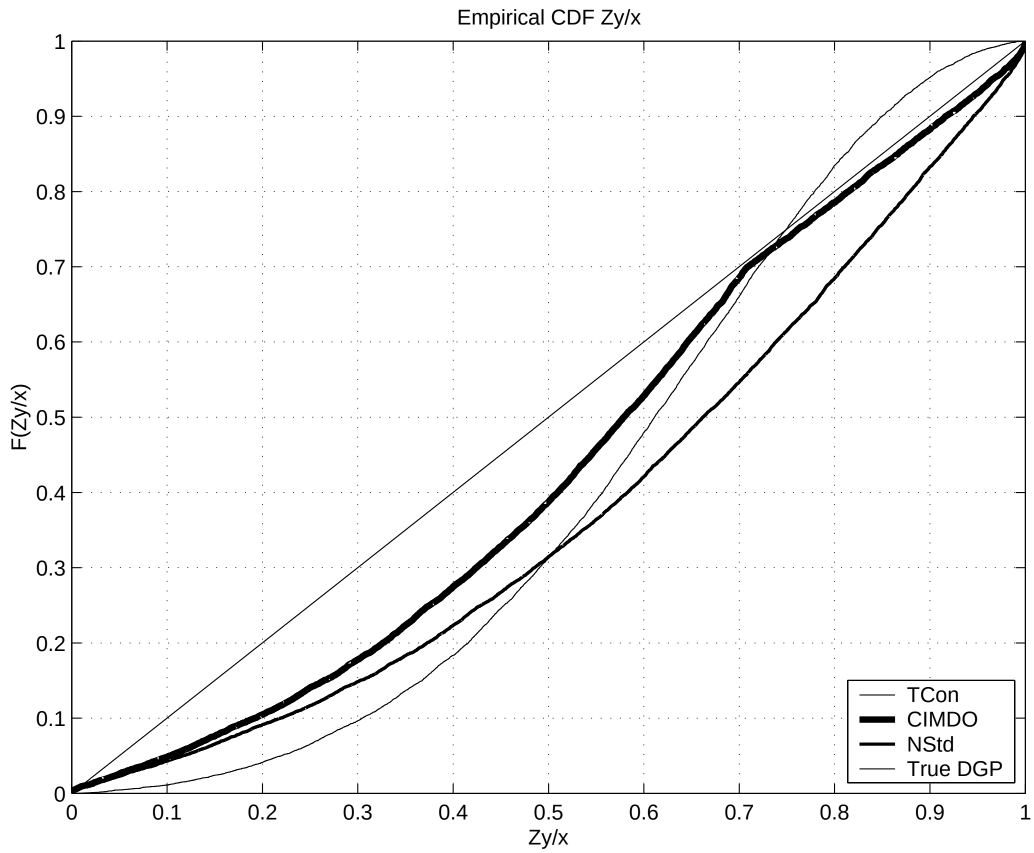

Figure 11: Empirical CDF Zy|x: CIMDO, NStd, TCon

In this figure, we compare the empirical cdf's of the $z_{y|x}$ variables corresponding to CIMDO, NStd, TCon and the true-DGP (45° degree line). It is clear that CIMDOs distribution outperforms NStd and TCon along the whole integration domain of the distribution. However, it is in the region of default (upper right corner) that the improvements are the best.

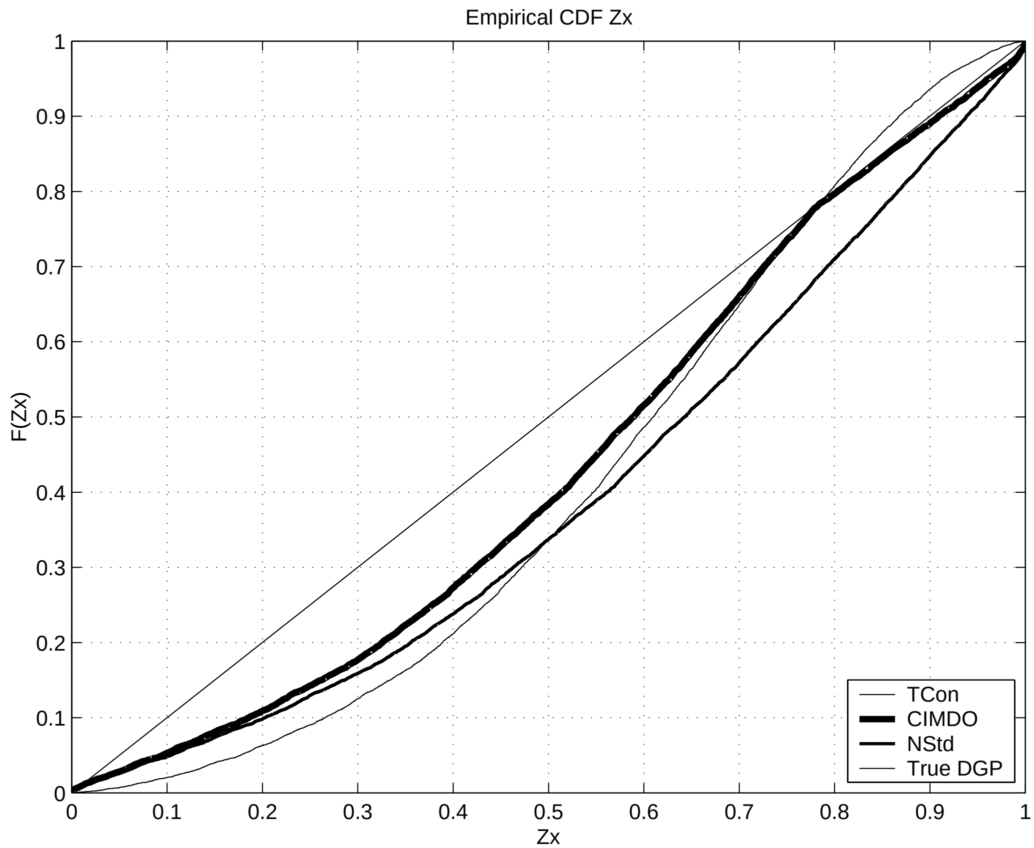

Figure 12: Empirical CDF Zx: CIMDO, NStd, TCon

In this figure, we compare the empirical cdf's of the z_x variables corresponding to CIMDO, NStd, TCon and the true-DGP (45° degree line). It is clear that CIMDOs distribution outperforms NStd and TCon along the whole integration domain of the distribution. However, it is in the region of default (upper right corner) that the improvements are the best.

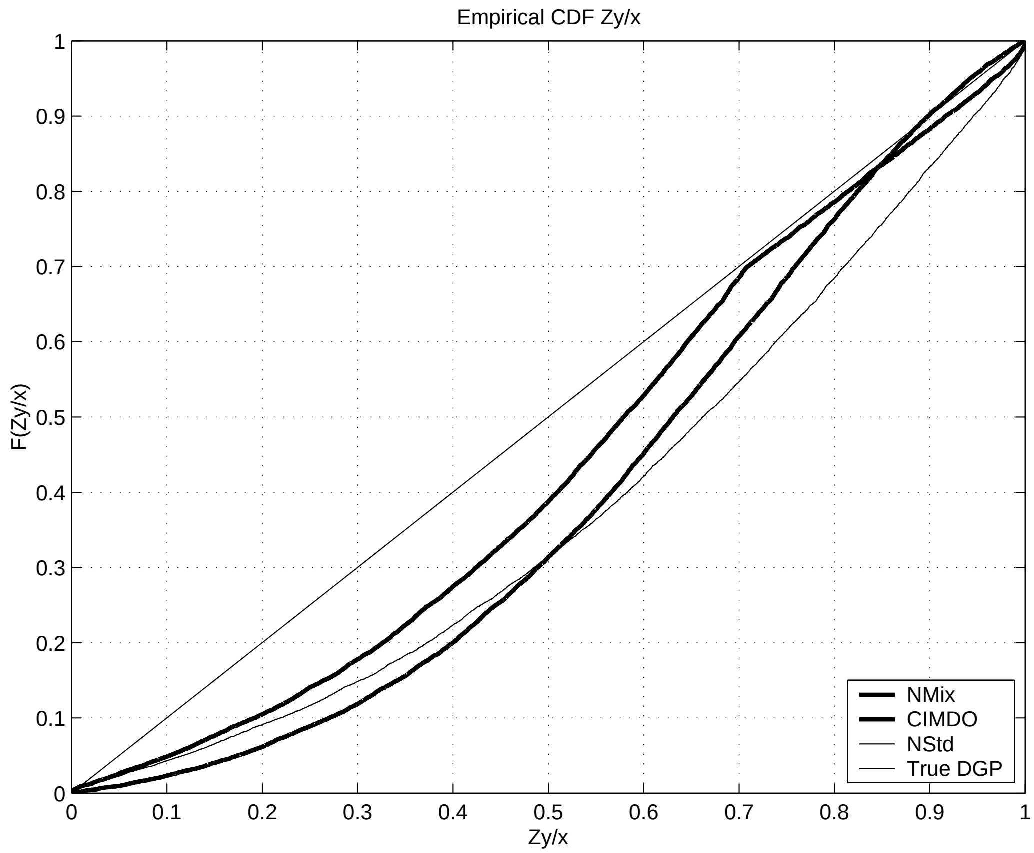

Figure 13: Empirical CDF Zy|x: CIMDO, NStd, NMix

In this figure, we compare the empirical cdf's of the $z_{y|x}$ variables corresponding to CIMDO, NStd, NMix and the true-DGP (45° degree line). It is clear that CIMDOs distribution outperforms NStd and NMix along the whole integration domain of the distribution. However, it is in the region of default (upper right corner) that the improvements are the best.

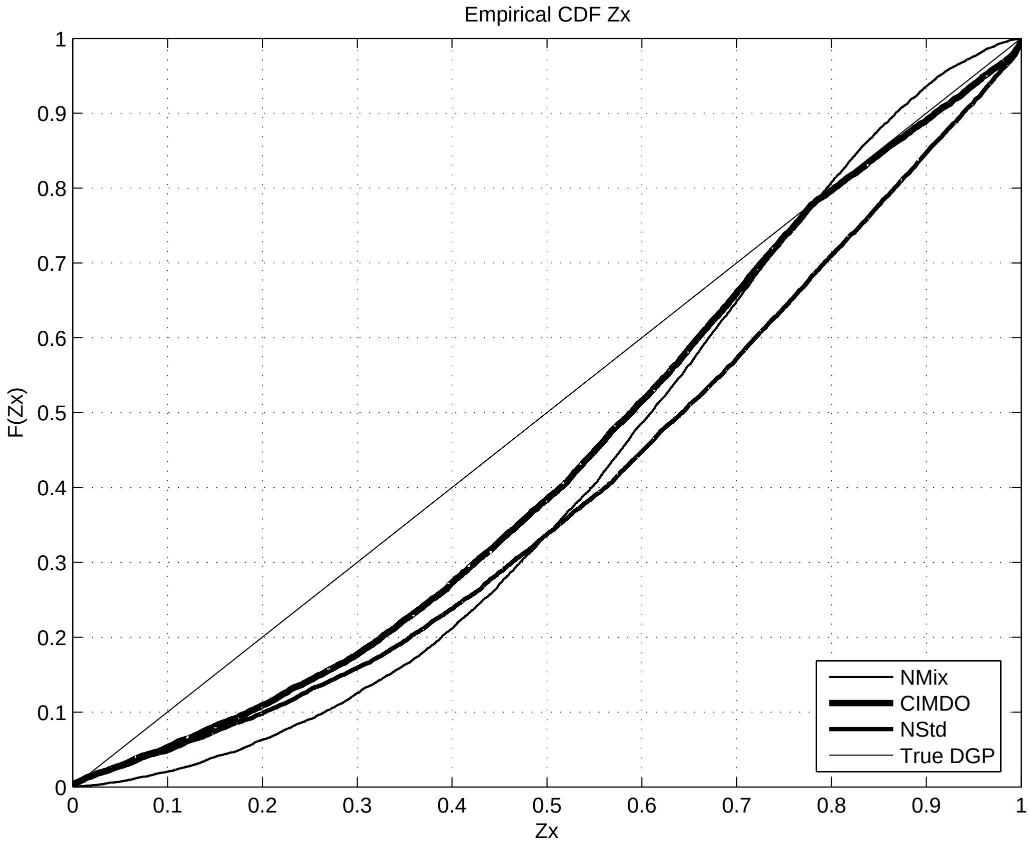

Figure 14: Empirical CDF Zx: CIMDO, NStd, NMix

In this figure, we compare the empirical cdf's of the z_x variables corresponding to CIMDO, NStd, NMix and the true-DGP (45° degree line). It is clear that CIMDOs distribution outperforms NStd and NMix along the whole integration domain of the distribution. However, it is in the region of default (upper right corner) that the improvements are the best.

6 CIMDO: density evaluation results under the PIT criterion

We present the empirical *cdf's* of the *z-variables* corresponding to CIMDO, NStd, NCon, TCon and NMix in Figures 3 to 14. From these figures, we can observe the following:

1. As we proved in Section 3, the *cdf's* of the PITs of the random variable realizations under the correct specification should always be represented by a 45° degree line, irrespective of knowing the true-DGP or not. Thus, the *cdf's* of the PITs of the random variables realizations, x and y under the specification that was assumed to represent the *true*-DGP, and that is indicated in equation (48), is a 45° degree line. This confirms that the realizations of the random variables x and y came from the assumed *true*-DGP.

   Therefore, the 45° *degree line can be taken as a benchmark to rank how good a particular distributional specification is, the closer to the* 45° *degree line, the better the specification.*

2. In Figures 3 to 14, it is clear that the CIMDO-recovered distribution outperforms the standard normal distribution under the PIT criterion. This is obviously not surprising, since the assumption that the multivariate distribution followed by the assets in a portfolio is the standard normal might appear to be naive and not consistent with the empirical facts. However, we make this comparison in order to get an idea of the degree of misspecification that could be reached when probabilities of default under the *prior* are completely inconsistent with the empirical probabilities of default.

3. When comparing the CIMDO-recovered distribution with the conditional normal distribution, the conditional t-distribution and the mixture model distribution, the CIMDO-derived distribution outperforms all of the competing distributions, including the mixture model which produces distributions that decay at a lower speed than the normal distribution (although it is still exponential). See that the improvement of the CIMDO-recovered distribution is evident along the whole domain of integration of the distribution.

4. Note though, that it is in the *region of default* of the CIMDO-derived distribution (upper right corner) that the improvement is best. This result is consistent with the fact that the information embedded in the CIMDO-derived density, via the *moment-consistency constraints,* relates to the *region of default.*

5. Above, we claimed that the specification of the CIMDO-derived density is closer to the correct specification than the specifications of the competing distributions along the whole domain of integration of the distribution, but especially in the region of default (upper right corner). However, when looking at the empirical *cdf's* of the *z-variables* it is apparent that outside the region of default, even the CIMDO-derived density is far from the 45° degree line. These discrepancies cause the Null hypothesis taken in the K-S test, H_0: $F = \mathbf{U}(\mathbf{0}, \mathbf{1})$, to be rejected for all the distribution specifications. This is not surprising because the K-S test considers the whole domain of integration of the distributions that are evaluated. Note, however, that our objective is to *rank the relative correctness* of their specifications. Given the data limitations at hand, we did not expect that either the distribution recovered by CIMDO, or the other distributions would be $\mathbf{U}(\mathbf{0}, \mathbf{1})$. Equivalently, the true distribution is unattainable given the information restrictions, therefore we focus on the relative improvements of the CIMDO-recovered distribution over the competing distributions.

6. Finally, notice that achievement of better accuracy in the upper right corner of the empirical *cdf* is important. This is because this is the range of the distribution's domain that has greatest relevance when computing the PLD of a portfolio of loans, since it represents, for each of the assets making up the portfolio, the probability of going into default.

   Data limitations frequently represent a challenge for the proper calibration of the parametric distributions that are usually employed for the modelling of the portfolio credit risk of SMEs and unlisted firms. In the most extreme cases, information restrictions do not allow any form of calibration. However, as we have already discussed, most of the time, although imperfectly, some sort of calibration is performed. The use of the empirically observed probabilities of default to "fit" parametric distributions is common practice[47]. The evaluation exercise presented in this chapter, although far from being exhaustive, provides preliminary evidence that CIMDO-derived densities outperform the conditional normal distribution, the conditional t-distribution and the mixture model distribution under the probability integral transformation criterion, when data restrictions do not allow their correct calibration. Nonetheless, studies using a more comprehensive number of assumed data generating processes driving the firms' underlying asset values and comparisons with a wider range of distributions are needed before one can make definite conclusions about the performance of CIMDO-derived densities.

7 Conclusions

Outstanding loans to non-publicly traded companies represent an important percentage of commercial banks' assets in developed and developing economies. However, the accurate modelling of the credit risk of the portfolios of these loans or loans given to arm's-length firms has been a major problem for financial institutions and regulators, due to the limited data available for their modelling. Current credit risk measurement methodologies incorporate convenient parametric assumptions in order to represent the unavailable information. But data restrictions may not allow for the proper calibration of the adopted assumptions and, as a result, significant uncertainty with respect to the correct distributional specifications may be introduced. Improvements in the methodologies used to measure the credit risk of portfolios of such assets can have important implications for banks' risk management and systems' financial stability.

CIMDO allows the recovery of multivariate distributions from the incomplete set of information available for the modelling of the credit risk of portfolios of loans granted to SMEs, unlisted and arm's-length firms; e.g., the frequencies of default of the marginal distributions of the loans making up the portfolio. This is possible without explicitly including information about the dependence structure between the assets comprising the portfolio. However, if such dependence structure information is available, it can be easily incorporated into the modelling framework. This is an important feature of the proposed methodology that makes its empirical implementation feasible and straightforward, because for portfolios of non-traded assets or arm's-length firms, dependence structure information is usually not accessible.

CIMDO is based on the cross-entropy formalism, which through an optimization framework, recovers distributions that are consistent with empirical observations, formulated as moment-consistency constraints. The entropy formalism seeks to make the "best" predictions possible from the information that is available and provides a basis for transforming the latter into a distribution of probabilities describing our state of knowledge.

Although far from being exhaustive, the exercise presented in Section 5.5 shows that CIMDO-recovered densities outperform the most commonly used parametric distributions for portfolio credit risk modelling of SMEs and unlisted firms, under the PIT criterion, when data restrictions do not allow their correct calibration.

The impact of economic shocks that affect the frequencies of default (which are used for the recovery of CIMDO-distributions) is embedded in CIMDO-inferred densities. As a result, the impact of economic shocks is passed onto the dependence structure embedded in such distributions. Therefore, CIMDO offers the possibility of improving the modelling of dependence through time; e.g. as business cycle conditions change. The CIMDO-copula modelling is not further developed in this paper; it is a line of research that constitutes a major project in itself, and one that we will be presenting in the near future.

[47] The use of proxy variables of listed or publicly-rated companies with similar credit risk classifications is some times practiced. However, using listed-firms' information as a proxy for non-listed firms is compeltely incorrect. Among theoretical considerations regarding continuous trading assumptions, different liquidity premiums apply to these markets. (Jarrow and Turnbull, 2000 and Longstaff et al, 2003).

References

[1] Altman, Edward, “Measuring corporate bond mortality and performance”, Journal of Finance, V.44, pp. 909-922, 1989.

[2] Avellaneda, M. (1998), “Minimum Relative Entropy Calibration of Asset Pricing Models”, International Journal of Theoretical and Applied Finance, Vol. 1, No. 4 (1998), p. 447-472.

[3] Basak, S., Saphiro, A. (2001), “A Model of Credit Risk, Optimal Policies, and Asset Prices”, London Business School and Stern School of Business, NYU.

[4] Bekaert G., Campbell R. (2002), “Emerging Markets Finance”, Columbia University.

[5] Blyth, S. (1996), “Out of Line”, Risk, 9(10), 82-84.

[6] Buche, P., Kelly, M. (1996), “The Maximum Entropy Distribution of an Asset Inferred from Option Prices”, Journal of Financial and Quantitative Analysis 31, p. 143-159.

[7] Butler, J., Schachter, B. (1998), “Estimating Value-at-Risk with a Precision Measure by Combining Kernel Estimation with Historical Simulation”, Review of Derivatives Research 1:371-390.

[8] Collin-Dufresne, P., Goldstein, R., Helwege, J. (2002), “Are Jumps in Corporate Bond Yields Priced? Modelling Contagion via the Updating of Beliefs”, Working Paper, Carnegie Mellon University.

[9] Cosella G., Berger, R. (1990), “Statistical Inference”, Duxbury Press, U.S.A.

[10] Crockett, A. (2000), “Marrying the Micro- and Macro-prudential Dimensions of Financial Stability”, BIS Review No 76.

[11] Crockett, A. (2002), Speech by A. Crockett at the Bankers Club Annual Banquet held in London, 4 February 2002.

[12] Crosbie, P., Bohn, J. (1998), “Modeling Default Risk”, KMV LLC.

[13] Crouhy, M., Mark, R., Galai, D. (2000), “Risk Management”, McGraw-Hill Trade.

[14] Danielsson, J. and de Vries, C.G. (1997), “Tail Index and Quantile Estimation with Very High Frequency Data”, JEF, volume 4, 241-257.

[15] Danielsson, J., Shin, H., Zigrand, J. (2001), “Asset Price Dynamics with Value-at-Risk Constrained Traders”, Discussion Paper 394, Financial Markets Group, London School of Economics.

[16] Danielsson, J., Zigrand, J. (2001), “What Happens When You Regulate Risk?”, Discussion paper 393, Financial Markets Group, London School of Economics.

[17] Diebold, F., Gunther, T., Tay, A, (1998), “Evaluating Density Forecasts with Applications to Financial Risk Management”, IER, 39, 863-883.

[18] Duffie, J. D., Singleton, K. J. (1999), “Modelling Term Structures of Defaultable Bonds”, Review of Financial Studies 12, 687-720.

[19] Embrechs, P., McNeil, A., and Straumann, D. (1999), “Correlation and Dependence Properties in Risk Management: Properties and Pitfalls”, Working Paper, RiskLab, ETH, Zurich.

[20] Embrechs, P., Resnick, S., Samorodnitsky, G. (1999), “Extreme Value Theory as a Risk Management Tool”, North American Actuarial Journal 26, 30-41.

[21] Embrechs P. (2000), “Extreme Value Theory: Potential and Limitations as an Integrated Risk Management Tool”, Working Paper, RiskLab, ETH, Zurich.

[22] Embrechs, P., McNeil, A., Frey, R. (2003), "Quantitative Methods for Financial Risk Management", Book Project, RiskLab, ETH, Zurich.

[23] Engle, R. (2003), "Dynamic Conditional Correlation and Tail Dependent Copulas", Stern School of Business, NYU.

[24] Fama, E.F. (1965), "The Behavior of Stock Market Prices", Journal of Business 38, 34-105.

[25] Fisher, R. A. (1932), "Statistical Methods for Research Workers".

[26] Frey, R., McNeil, A.J. (2001), "Modelling Dependent Defaults". Working Paper, Department of Mathematics, ETH Zurich.

[27] Gagliardini, P., Gorieroux, C. (2003), "Spread Term Structure and Default Correlation", Universita della Svizzera Italiana.

[28] Gilbarg, D., Trudinger, N. (1977), "Elliptic Partial Differential Equations of Second Order", Springer-Verlag, 2nd ed, p. 147.

[29] Glasserman, P., Heidelberger, P., Shahabuddin, P. (2000), "Portfolio Value-at-Risk with Heavy-Tailed Risk Factors", working paper, Money, Economics and Finance, Columbia Business School, Columbia University.

[30] Gokhale, D., Kullback, S. (1978), "The information in contingency tables", New York, Marcel Dekker.

[31] Good, I. (1963), "Maximum Entropy for Hypothesis Formulation, Especially for Multidimensional Contingency Tables", Annals of Mathematical Statistics 34, p. 911-34.

[32] Gordy, M. (2000), "A Comparative Anatomy of Credit Risk Models", Journal of Banking and Finance 24, p. 119-149.

[33] Gupton, M., Finger, C., Bhatia, M. (1997), "CreditMetrics, Technical Document", Morgan Guaranty Trust Company, Risk Management Research.

[34] Hansen, B. (1994), "Autoregressive Conditional Density Estimation", International Economic Review 35, 705-730.

[35] Hoskin, J.R.M., Bonti, G., and Siegel, D. (2000), "Beyond the Lognormal", Risk 13, No. 5 (May), 59-62.

[36] Jarrow, R. A., Turnbull, S. M., (1997), "When Swaps are Dropped", Risk, Vol 10, N5.

[37] Jaynes, E. (1957), "Information Theory and Statistical Mechanics", Physics Review 106, p. 620-30.

[38] Jaynes, E. (1957b), "Information Theory and Statistical Mechanics II", Physics Review 108, p. 171-90.

[39] Johnson, N., Kotz, S. (1970), "Distributions in Statistics: Continuous Univariate Distributions", John Wiley and Sons.

[40] Koyluoglu, U., Wilson, T., Yague, M. (2003), "The Eternal Challenge", Working paper in progress, Mercer Oliver Wyman.

[41] Kullback, J., Leibler, R. (1951), "On Information and Sufficiency", Annals of Mathematical Statistics, 22, 79-86.

[42] Kullback, J., (1959), "Information Theory and Statistics", John Wiley, New York.

[43] Lando, D. (1998), "On Cox Processes and Credit Risk Securities", Review of Derivatives Research, 2, 99-120.

[44] Lee, T., Judge, G. (1996), "Entropy and Cross Entropy Procedures for Estimating Transition Probabilities from Aggregate Data", Bayesian Conference in Statistics and Econometrics: Essays in Honor of Arnold Zellner. John Wiley, New York.

[45] Levine, R., Tribus, M. (1979), "The Maximum Entropy Formalism", MIT Press, Cambridge, MA.

[46] Lopez, J., Walter, C. (2000), "Evaluating Covariance Matrix Forecasts in Value-at-Risk Framework", FRBSF Working Paper 2000-21, Federal Reserve Bank of San Francisco.

[47] McLachlan, G., Basford, K. (1987), "Mixture Models: Inference and Applications to Clustering, Marcel Dekker", New York and Basel.

[48] McNeil, A., Frey, R. (1999), "Estimation of Tail-related Risk Measures for Heteroscedastic Financial Time Series: an Extreme Value Approach", Working Paper, RiskLab, ETH, Zurich.

[49] Merton, R. (1974), "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates", The Journal of Finance, Vol 29, p. 449-470.

[50] Mina, J., Xiao, Y. (2001), "Return to RiskMetrics: The Evolution of a Standard", RiskMetrics.

[51] Morris, S., Shin, H. (1999), Risk Management with Interdependent Choice, Oxford Review of Economic Policy 15-3 (1999), 52-62.

[52] Nelsen, R. (1999), "An Introduction to Copulas", Lecture Notes in Statistics 139, Springer.

[53] Nickell, P., Perraudin, W., Varotto, S. (2000), "Stability of Rating Transitions", Journal of Banking and Finance, 24, 203-227.

[54] Patton, A. J. (2001a), "Modelling Time-Varying Exchange Rate Dependence Using the Conditional Copula", University of California, San Diego.

[55] Patton, A. J. (2001b), "Estimation of Copula Models for Time Series of Possibly Different Lengths", Working Paper, University of California, San Diego.

[56] Rosenblatt, M. (1952), "Remarks on a Multivariate Transformation", Annals of Mathematical Statistics, 23, 470-472.

[57] Saunders, A., Allen, L. (2002), "Credit Risk Measurement, New Approaches to Value at Risk and Other Paradigms", 2. edition, Wiley Finance Editions.

[58] Schonbucher, P., Schubert, D. (2001), "Copula-dependent Default Risk in Intensity Models", Working paper, Bonn University.

[59] Schonbucher, P. (2003), "Credit Derivatives Pricing Models", John Wiley & Sons, Ltd. U.K.

[60] Shannon, C. (1948), "A Mathematical Theory of Communication", Bell System, Technical Journal 27, p. 379-423.

[61] Shaw, J. (1999), "Beyond VAR and Stress Testing, 231-244, in Monte Carlo: Methodologies and Applications for Pricing and Risk Management", B. Dupire, ed., Risk Publications, London.

[62] Servigny, A., Renault,O. (2002), "Default Correlation: Empirical Evidence", S&P Risk Solutions.

[63] Turnbull, S. (2003), "Practical Issues in Modelling Default Dependence", Working Paper in Progress, Bauer College of Business, University of Houston.

[64] Yu, F. (2003), "Correlated Defaults in Reduced Form Models", Working Paper in Progress, University of California, Irvine.

[65] Zangari, P. (1996), "An Improved Methodology for Measuring VaR", RiskMetrics Monitor, 2nd quarter pp.7-25.

[66] Zellner, A. (1988), "Optimal Information Processing and Bayes Theorem", American Statistician 42, p. 278-84.

Appendix

A.1. The entropy function

Using the entropy concept developed in the *XIXth* century by Boltzman and continued lately by Maxwell, Gibbs, Bernoulli, Laplace, Shannon (1948) developed the "entropy of the distribution of probabilities" to measure the uncertainty of a collection of events.

In developing this approach, Shannon (1948) supposed that an experiment with N trials (repetitions) was carried out. This experiment had K possible outcomes (states). He assumed that $N_1, N_2, ..., N_K$ represented the number of times that each outcome occurs in the experiment of length N, where $\sum N_k = N$, $N_k \geqslant 0$, and $k = 1, 2, ..., K$.

In this setting there are N trials and each trial has K possible outcomes; therefore, there are K^N conceivable outcomes in the sequence of N trials. Of these, a particular set of frequencies

$$p_k = \frac{Nk}{N} \qquad \text{or} \qquad N_k = Np_k \qquad \text{for} \qquad k = 1, 2, ..., K,$$

can be realized in a given number of ways as measured by the multiplicity factor (possible permutations). Thus, the number of ways that a particular set of N_k is realized, can be represented by the multinomial coefficient

$$W = \frac{N!}{Np_1!Np_2!....Np_k!} = \frac{N!}{\prod_k N_k!},$$

or its monotonic function

$$lnW = lnN! - \sum_{k=1}^{K} \ln N_k! \tag{A.1.1}$$

Stirling's approximation $lnx! \approx x \ln x - x$ as $0 < x \to \infty$ is used to approximate each component on the right hand side of (A.1.1). Then, for large N, we have

$$lnW \approx N \ln N - N - \sum_{k=1}^{K} \ln N_k \ln N_k + \sum_{k=1}^{K} N_k$$

since $\sum_{k=1}^{K} N_k = N$, we get

$$lnW \approx N \ln N - \sum_{k=1}^{K} N_k \ln N_k. \tag{A.1.2}$$

The ratio $\frac{N_k}{N}$ represents the frequency of the occurrence of the possible K outcomes in a sequence of length N and $\frac{N_k}{N} \to p_k$ as $N \to \infty$. Consequently (A.1.2) yields,

$$\begin{aligned} lnW &\approx N \ln N - \sum_{k=1}^{K} Np_k \ln (Np_k) \\ &= N \ln N - \sum_{k=1}^{K} N_k \ln N - N \sum_{k=1}^{K} p_k \ln p_k \\ &= -N \sum_{k=1}^{K} p_k \ln p_k. \end{aligned}$$

Finally,

$$\begin{aligned} N^{-1} lnW &\approx -\sum_{k=1}^{K} p_k \ln p_k \\ H(p) &= -\sum_{k=1}^{K} p_k \ln p_k \end{aligned}$$

Which is the Shannon entropy measure, where $p_k \ln p_k = 0$ for $p_k = 0$.

A.2. The prior distribution

In this paper, the *posterior* **p** and *prior* **q** distributions, to which we refer to in the Minimum Cross Entropy Distribution MXED, are different from the Bayesian context. This is because the *posterior* **p** in the MXED is obtained via an optimization rule. However, we use this terminology because it is the standard terminology in the entropy literature. This might be the case because although not the same, there is a close connection between the two efficient information processing rules.

As noted by Lee and Judge (1996), in Bayes' rule, given initial information I, there are two inputs and two outputs. The two inputs are $\pi(\theta|I)$, the *prior* density for the parameter θ, and the data density $f(y|\theta, I)$. The two outputs are the post data *posterior* density $g(\theta|y, I)$ and $h(y|I)$ the marginal density for y. In terms of these pieces, Bayes' rule to transform prior and sample information into posterior information is:

$$g(\theta/y, I) = \frac{\pi(\theta|I) f(y|\theta, I)}{h(y|I)}. \tag{A.2.1}$$

Contrasting (A.2.1) and (14), we note in (14) that $\widehat{p(x)}$ corresponds to the posterior $g(\theta|y, I)$, that $q(x)$ corresponds to the prior $\pi(\theta|I)$, that the factor $\exp\left[\sum_{t=1}^{T} \widehat{\lambda}_t f_t(x)\right]$ corresponds to the data density $f(y|\theta, I)$ and the numerator of equation (14) corresponds to the marginal density $h(y|I)$, that serves to convert the relative probabilities into absolute ones.

A.3. Definitions

Definition A.3.1: Quasi-inverse of a distribution function

Given a non-decreasing function G: $\mathbb{R} \to \mathbb{R}$. The generalized inverse of G is defined by:

$$G^{\leftarrow}(y) = \inf\{x \in \mathbb{R} : G(x) \geq y\}. \tag{A.3.1}$$

Where the convention that the infimum of an empty set is ∞. Note that $G^{\leftarrow}(y) = G^{-1}(y)$, the usual inverse of G, when G is strictly increasing.

Definition A.3.2: Partial Moment of a Distribution Function

Partial moments are measures of risk based on the lower or upper part of a distribution. Given an exponent $r \geq 0$ and a reference point Z, the upper partial moment $UPM(r, Z)$ is defined as:

$$UPM(r, Z) = \int_{Z}^{\infty} (x - Z)^r dF_X(x), \tag{A.3.2}$$

for a r.v. x with cumulative density function (*cdf*) F_X.

A.4. K-S test results

We computed the Kolmogorov-Smirnov (K-S) tests with H_0: $F = \mathbf{U(0,1)}$, H_a: $F \neq \mathbf{U(0,1)}$ at the 5% significance level. H=1 if H_0 is rejected. K-S tests results are as follows:

Table 5: K-S test: Zy|x

K-S test: $z_{y/x}$					
	CIMDO	**NStd**	**NCon**	**TCon**	**NMix**
H	1	1	1	1	1
K-Statistic	0.1281	0.1875	0.2246	0.2179	0.2003
Critical Value	0.0136	0.0136	0.0136	0.0136	0.0136

Table 6: K-S test: Zx

K-S Test: z_x					
	CIMDO	**NStd**	**NCon**	**TCon**	**NMix**
H	1	1	1	1	1
K-Statistic	0.1301	0.1664	0.1933	0.1902	0.1854
Critical Value	0.0136	0.0136	0.0136	0.0136	0.0136

Table 7: K-S test: Zx|y

K-S Test: $z_{x/y}$					
	CIMDO	**NStd**	**NCon**	**TCon**	**NMix**
H	1	1	1	1	1
K-Statistic	0.1296	0.1654	0.1932	0.1834	0.1700
Critical Value	0.0136	0.0136	0.0136	0.0136	0.0136

Table 8: K-S test: Zy

K-S test: z_y					
	CIMDO	**NStd**	**NCon**	**TCon**	**NMix**
H	1	1	1	1	1
K-Statistic	0.1287	0.1883	0.2251	0.2237	0.2218
Critical Value	0.0136	0.0136	0.0136	0.0136	0.0136