On the Way to Recovery: A Nonparametric Bias Free Estimation of Recovery Rate Densities

Olivier RENAULT[a] and Olivier SCAILLET[b] [1]

[a] Standard & Poor's Risk Solutions, 18 Finsbury Circus, London EC2M 7NJ, United Kingdom.
[b] HEC Genève and FAME, UNIMAIL, Faculté des SES, 102 Bd Carl Vogt, 1211 Genève 4, Switzerland.

First version: July 2002
This version: September 2003

Abstract

In this paper we analyse recovery rates on defaulted bonds using the Standard and Poor's / PMD database for the years 1981-1999. Due to the specific nature of the data (observations lie within 0 and 1), we must rely on nonstandard econometric techniques. The recovery rate density is estimated nonparametrically using a beta kernel method. This method is free of boundary bias, and Monte Carlo comparison with competing nonparametric estimators show that the beta kernel density estimator is particularly well suited for density estimation on the unit interval. We challenge the usual market practice to model parametrically recovery rates using a beta distribution calibrated on the empirical mean and variance. This assumption is unable to replicate multimodal distributions or concentration of data at total recovery and total loss. We evaluate the impact of choosing the beta distribution on the estimation of credit Value-at-Risk.

Key words : default, recovery, kernel estimation, credit risk.

JEL Classification : C13, C14, C51, G18, G21, G33.

[1] We thank the editor G. Szegö as well as two referees for their helpful comments. The second author gratefully acknowledges financial support from the Swiss National Science Foundation through the National Center of Competence: Financial Valuation and Risk Management (NCCR FINRISK). Part of this research was done when the second author was visiting THEMA and IRES. The views expressed therein are those of the authors but not necessarily those of Standard and Poor's.

1 Introduction

Regulatory reform and in particular the revised Basel Accord (see the (2003) consultative document) has led to renewed interest in the measurement and modelling of key inputs of credit risk portfolio models: probabilities of default, recovery rates and correlations. This has coincided with the explosion of the credit derivatives market (see Duffie and Singleton (2003) for a survey) which rely on similar inputs. Until recently most empirical research on default data has focussed on modelling and estimating default probabilities, either from ratings, macroeconomic factors or equity prices.

The second key building block of portfolio models is the recovery in default or loss given default (LGD, defined as one minus the recovery rate) function, typically expressed in terms of a ratio (dollar recovery / amount invested). This crucial element has been relatively unexplored and recovery analysis has concentrated on comparisons of the mean and the variance of recovery rates for various defaulted issues with different characteristics (see however Hu and Perraudin (2002) for a study of marginal impact of particular bond characteristics by means of multivariate regressions). Altman and Kishore (1996) for example carry out a detailed analysis of recovery rates per industry and seniority levels while rating agencies provide regular updates on the recovery experience (see e.g. Keisman and Van de Castle (1999), Bahar and Brand (2000)). A thorough study of bond recovery rates can be found in Altman et al. (2001), with some interesting insight into the link between default rates and recoveries.

Mean and variance however are insufficient to fully specify the recovery function. In order to put more structure on the model it has become market practice to calibrate a beta distribution to the first two empirical moments. In this paper we analyse recovery functions nonparametrically. By doing so we achieve two things: first, we are able to visualize the recovery function without assuming a specific parametric specification. Second, we can assess the appropriateness of the beta distribution and the potential bias introduced by this standard parametric assumption on portfolio losses as measured by credit Value-at-Risk (VaR) for example.

The paper is organised as follows. Section 2 outlines our estimation strategy. We use a nonparametric density estimator based on the asymmetric beta kernel introduced by Chen (1999). We choose this estimator because it has been especially designed for observations lying in $[0, 1]$ and has nice properties such as absence of estimation bias at the boundaries. This property is particularly relevant here since 0 and 100% recovery rates are of primary importance in credit risk management. We favour a kernel density estimator over the standard histogram since a smoothed estimator avoids the unattractive feature of estimating the unknown recovery densities by a step function. Clearly smoothness eases comparison among nonparametric and parametric estimates. Besides the sensitivity of the histogram (see Wand and Jones (1985) p. 6 for a convincing illustrative example and further discussion) to the placement of the bin edges is a problem not shared by the kernel density estimator. From a theoretical point of view, the mean integrated squared error (MISE) of the histogram is also

known to be asymptotically inferior to the kernel density estimator since its convergence rate (Scott (1979)) is lower than the one of the kernel estimator. This means that the histogram does not use the data as efficiently as the kernel estimator. In Section 3 we report Monte Carlo results concerning the finite sample properties of the beta kernel density estimator. A performance comparison with two other nonparametric density estimators is provided. Section 4 contains the nonparametric study of the recovery functions per seniority level and per industry. The analysis exploits recovery rates on defaulted bonds extracted from the Standard and Poor's / PMD database for the years 1981-1999. In Section 5, we investigate the practical consequence of modelling adequately the recovery function in credit risk management. In particular we confront credit VaR obtained from the market practice based on a beta distribution specification and our nonparametric approach. Section 6 concludes.

2 The beta kernel density estimator

The most popular nonparametric estimator of an unknown probability density function is the standard Gaussian kernel estimator (see e.g. Silverman (1986), Haerdle and Linton (1994) or Pagan and Ullah (1999) for an introduction). Its consistency is well documented when the support of the underlying density is unbounded, i.e. when data live on $(-\infty, +\infty)$. This estimator is however no more appropriate in the case of a bounded support. We know that there exists a boundary bias as the symmetric Gaussian kernel will assign non zero probability outside the support of the distribution when smoothing is carried out near a boundary. In the case of our study, it would imply that negative recoveries are possible.

Recently, Chen (1999) has proposed a beta kernel density estimator for densities defined on $[0, 1]$. This estimator is based on the asymmetric beta kernel which exhibits two special appealing properties: a flexible form and a location on the unit interval. The kernel shape is allowed to vary according to the data points, thus changing the degree of smoothing in a natural way, and its support matches the support of the probability density function to be estimated. This leads to larger effective sample sizes used in the density estimation and usually produces density estimates that have smaller finite-sample variances than other nonparametric estimators. The beta kernel density estimator is simple to implement, free of boundary bias, always non negative, and achieves the optimal rate of convergence ($n^{-4/5}$) for the mean integrated squared error (MISE) within the class of nonnegative kernel density estimators (see Chen (1999) for details). Furthermore, even if the true density is infinite at a boundary, the beta kernel estimator remains consistent (Bouezmarni and Rolin (2001)). This unboundeness can arise empirically from a clustering of observations near boundaries. This consistency property is thus particularly important for the study of recovery rates as 0% and 100% recoveries are frequent events. Similar asymmetric kernel estimators for densities defined on $[0, +\infty)$ have been studied in Chen (2000), Scaillet (2001), and Bouezmarni and Scaillet

(2003). These estimators share the same valuable properties than those of the beta kernel estimator used in this paper.

Let $X_1, ..., X_n$ be a random sample from a distribution with an unknown probability density function f on the unit interval. In our case this sample will be made of observed recovery rates between 0% and 100%. Then, the beta kernel estimator of f at point x in $[0, 1]$ is formally defined as:

$$\hat{f}(x) = \frac{1}{n} \sum_{t=1}^{n} K(X_i, x/b + 1, (1 - x)/b + 1), \tag{1}$$

where the asymmetric kernel $K(.)$ is the beta probability density function:

$$K(u, \alpha, \beta) = \frac{1}{B(\alpha, \beta)} u^{\alpha-1}(1 - u)^{\beta-1}, \qquad u \in [0, 1],$$

with $B(.)$ denoting the beta function and b being a smoothing parameter, called the bandwidth, such that $b \to 0$ as $n \to \infty$.

It can be noticed that this estimator is easy to implement since it corresponds to straight averaging of a known function of the observed data points. Basically it simply requires to substitute the beta kernel for the Gaussian kernel in the standard symmetric kernel estimator.

3 Finite sample performance

In this section, we investigate the finite sample properties of the beta kernel density estimator on sets of data drawn from known distributions with $[0, 1]$ support.

For the sake of comparison we also compute estimates given by the standard Gaussian kernel density estimator. The direct use of the Gaussian kernel density estimator on data taking values in $[0, 1]$ will lead to a boundary bias when smoothing is carried out near a boundary, and we may expect a poor performance of this estimator for observations lying in the unit interval.

The Gaussian kernel estimator however does exhibit nice properties in the usual case of observations on $(-\infty, +\infty)$. This suggests a way to go around the boundedness problem using a three step procedure based on i) a monotonic mapping of the data from $[0, 1]$ to $(-\infty, +\infty)$, ii) a Gaussian kernel estimation of the density of the transformed data, and iii) transforming back the estimated density by multiplying it with the derivative of the inverse mapping. Such an estimator is called a transformation kernel density estimator (see Devroye and Gyoerfy (1985) and Silverman (1986) for early proposals). In this paper we have chosen to work with the logistic function because of the tractability of this mapping and because the logistic function is familiar in credit risk applications (in particular in credit scoring). The transformed data Y_i are then equal to $\ln(X_i/(1-X_i))$. Note that an undesirable feature of this type of estimator is that the best choice of the transformation depends quite heavily on the (unknown) shape of the true density f (see e.g. Wand and Jones (1995) Section 2.10.3 for

a discussion). We also examine below the finite sample properties of this third estimator.

In the Monte Carlo design we consider four distributions simulated using the beta distribution with different parameters (Figure 1). Table 1 reports the parameters for each simulation run. The first distribution is symmetric w.r.t. 0.5 and has zero probability mass at endpoints (bell shaped). We can thus anticipate that the standard symmetric estimator will perform well in that case because the boundary bias is small in absolute terms, and most observations are located around the center of the distribution. This can be seen as a worse case scenario for the performance of the other two estimators: the beta kernel estimator and the Gaussian estimator on transformed data. This configuration of recoveries is unlikely to arise in practice because recovery densities tend to be skewed and have a concentration of observations near the boundaries (see next sections). The three other cases are closer to real life situations. The U-shaped density has unbounded densities at both endpoints, and can be taken as a kind of "bang-bang" case: many 0% and 100% recovery rates, and few in between. The final two cases exhibit skews towards 0% and 100% recovery rates, respectively. We thus have four specifications: bell-shaped, U-shaped, skewed left and skewed right.

The experiments are based on 1000 random samples of length $n = 100$ and $n = 1000$. The sample size $n = 100$ is close to the number of observations encountered in practice for subgroups of recovery rates, for example by industry sector. For each simulated sample and each estimator, integrated squared errors (ISE): $\int(\hat{f} - f)^2$ were computed optimally from a grid of bandwidth values proportional to $n^{-2/5}$ for the asymmetric beta kernel and proportional to $n^{-1/5}$ for the symmetric Gaussian kernel. Optimal bandwidths (in terms of the mean integrated squared error) for estimation of densities are proportional to these powers of the sample size. Numerical integration was performed by a Gauss Legendre quadrature with 96 knots.

Average ISE are reported in Table 1. The beta kernel density estimator performs better than the standard Gaussian kernel estimator for the U-shaped, skewed left and skewed right distributions. It has however a higher average ISE for the bell-shaped distribution. The transformation kernel density estimator is less precise than the beta kernel estimator in each case and less precise than the Gaussian kernel estimator for the bell-shaped distribution. The standard estimator performs better, but only slightly with respect to the beta kernel estimator, for the first distribution only as expected. In view of these Monte Carlo results and the fact that they prevent assignment of non zero probability outside the support $[0, 1]$, we may conclude that beta kernels should be well suited for estimating recovery functions. This is done in the next section.

Table 1: Average ISE computed on 1000 samples.

$n = 100$	mean	var	α	β	ISE beta	ISE Gauss	ISE Trans
Bell-shaped	0.5	0.04	2.6250	2.6250	0.0309	0.0233	0.0329
U-shaped	0.5	0.16	0.2812	0.2812	0.0469	0.1782	0.0499
Skewed left	0.7	0.04	2.9750	1.2750	0.0268	0.0442	0.0502
Skewed right	0.3	0.04	1.2750	2.9750	0.0276	0.0513	0.0432
$n = 1000$	mean	var	α	β	ISE beta	ISE Gauss	ISE Trans
Bell-shaped	0.5	0.04	2.6250	2.6250	0.0056	0.0047	0.0061
U-shaped	0.5	0.16	0.2812	0.2812	0.0046	0.0073	0.0071
Skewed left	0.7	0.04	2.9750	1.2750	0.0056	0.0148	0.0091
Skewed right	0.3	0.04	1.2750	2.9750	0.0056	0.0149	0.0091

INSERT FIGURE 1 HERE

4 Empirical analysis of recovery rates

In this section we first discuss general considerations about recovery rates. We then describe the recovery data before presenting estimated recovery functions per seniority level and per industry via a beta kernel approach.

4.1 Recovery rates

Recovery rates are arguably more difficult to grasp than probabilities of default. There are two main measures of recovery given default: recoveries shortly after the default event based on trading prices and ultimate (discounted) recoveries. The first measure, used in this paper, relies on market data: recovery rates are the prices of defaulted securities (bonds and loans) in the distressed debt market recorded shortly after the default event. Ultimate recoveries are values of cash and assets actually recovered when the default process is settled, discounted back to the date of default. Trading price recoveries have several advantages: they are easy to measure (one just needs to observe a transaction price), they do not require the choice of a specific discount rate, and they correspond to the price an investor would recover if he were to liquidate his position immediately after default. They also have some significant drawbacks that are important to mention before we proceed to the estimation stage. Trading price recoveries are not available for all defaulted bonds and loans: they rely on the availability of a distressed debt market. They are also very sensitive to supply and demand issues: vulture funds have limited capacity and may not be able to absorb very large quantities of defaulted securities in a short period of time. Finally, trading price recoveries tend to exhibit significantly lower means than ultimate recoveries. They can therefore be misleading for a bank going through the complete workout process.

These differences are important to bear in mind. In practice, both measures are useful and the choice of a specific measure should correspond to the policy

adopted by the investor: either ultimate recoveries if the investor decides to hold on his securities until emergence from default or trading price recoveries if he chooses to liquidate his distressed lines soon after default. The above differences in definitions translate into differences in data features as well. Trading price recoveries are overwhelmingly in the range $[0, 1]$ with rare occurences slightly above 100%. Ultimate recoveries exhibit many more data points above 100% and even some reaching 120% (see Friedman and Sandow (2003)) and point masses at 0% and 100% are much heavier using this measure.

In what follows, we consider trading price recoveries only.

4.2 Recovery data

The data we use in this paper have been extracted from the S&P / PMD database. The data set comprises 623 defaulted bond issues from 1981 to 1999. Each issue is classified by seniority (senior secured, senior unsecured, subordinated and junior subordinated) and by industry (for S&P industry classification, see Figure 2).

The methodology used to calculate the recoveries is explained in details in Bahar and Brand (2000). The slight difference between their descriptive statistics and those shown below (Table 1) is explained by the fact that they include some non-US obligors while we restrict ourselves to American corporates only.

Table 2: Recoveries per seniority level.

Seniority	Number of observations	Mean recovery (%)	Standard deviation (%)
Senior secured	82	56.31	23.61
Senior unsecured	225	46.74	25.57
Subordinated	174	35.35	24.64
Junior subordinated	142	35.03	22.09
Total	623	42.15	25.42

Tables 2 and 3 report the number of observations, and the mean and standard deviation of the recovery rate per seniority and industry respectively. These results are in line with other studies on recovery such as Altman and Kishore (1996) or Hu and Perraudin (2002). As expected, the more senior the bond, the higher the recovery rate. Sectors with a lot of real assets such as the utility sector exhibit the highest mean recovery rate while Telecom bond investors tend to suffer the highest losses in default. Note however that our data does not include the recent wave of defaults in that industry. There does not seem to be significant differences in the volatility of recoveries by seniority but it tends to fluctuate wildly per industry.

Table 3: Recoveries per industry.

Industry*	Number of observations	Mean recovery (%)	Standard deviation (%)
Utility	55	68.37	20.82
Insurance	33	39.79	26.70
Telecom	14	24.73	7.53
Transport	39	37.52	27.12
Financial	24	29.70	24.63
Chemicals	22	36.51	26.33
High tech	14	50.46	22.15
Automotive	79	42.98	21.36
Building	43	39.69	29.46
Consumer	155	36.80	21.21
Leisure	86	42.91	27.06
Energy	59	45.56	25.61

*the complete industry name is provided in Figure 2.

4.3 Estimating the recovery function nonparametrically

Figure 2 shows the nonparametric estimates of the recovery rate densities based on the beta kernel approach. Calibrated beta densities based on market practice (see below) are also reported on the same graphs. Bandwidths are selected according to the usual rule of thumb, which consists in multiplying powers of the sample size ($n^{-2/5}$) and empirical standard deviation of observed data to get the bandwidth value.

Nonparametric plots of the recovery function usually exhibit several local modes. Bimodal distributions have also been reported in Asarnow and Edwards (1995) and Schuermann (2003). They appear to be approximately bell-shaped for the telecom sector (Industry 2) but are almost flat for the forest / building sector (Industry 9) and strongly skewed left for the utility sector (Industry 7) for example. It can be further noticed that nonparametric density estimates are always strictly positive and finite at the boundaries. This contrasts with beta densities which can only take the values 0 or $+\infty$ at the boundaries by definition. Recall that estimates for 0% and 100% recovery rates are particularly important for credit risk management purposes as they correspond to complete recovery and total loss, respectively. The case of the high tech sector (Industry 6) is particularly interesting. The beta distribution is unable to capture at the same time the bell shape in the center of the distribution and the large number of very low and very high recoveries. It assigns a zero density at 0% and 100% while the kernel appropriately captures that feature of the data.

Looking at Figure 2.a one can observe the intuitive fact that the recovery distribution is skewed left for the most senior bonds and skewed right for junior debt.

INSERT FIGURE 2 HERE

We have checked (results not reported here) on several examples that the shape of the estimated densities remains unaffected when 10% of the data are deleted randomly in the sample. This should not come as a surprise since we rely on a nonparametric approach based on a smoothing of data points. This smoothing process clearly limits the influence of potential outliers, and is part of the robustness properties of asymmetric kernel methods.

5 Implications for credit risk management

Over the past few years, it has become market practice to use a beta distribution to model recovery rates. A number of industry-sponsored models are specified as such (e.g. PortfolioManager and CreditMetrics). As we have seen above, the pattern of the recovery rate distribution can vary very significantly across seniority level and industries and it may be questionable to use a distribution calibrated on mean and variance only. In this section, we first take a closer look at the way practitioners typically calibrate their loss rate model. We then compare credit VaR derived from the use of beta distributed recovery rates to those obtained using directly empirical recovery rates.

5.1 Calibrating the recovery function

Many portfolio credit risk models rely on a beta distribution as recovery or loss function. For example, Credit Metrics relies on Monte Carlo simulations to obtain the number of defaults in a portfolio and each default is associated to a random draw in a beta distribution (see JP Morgan (1997), Crouhy, Galai and Mark (2001) or De Servigny and Renault (2003b)). This distribution is typically calibrated on the historical mean and variance of the losses, either calculated internally by banks or reported in default recovery studies such as Keisman and Van de Castle (1999). This method-of-moment approach, based on the historical mean and variance, induces a loss of efficiency. A Maximum Likelihood (ML) approach is therefore more appropriate if one believes that recovery rates are indeed beta distributed. A direct algebraic solution of the ML equations cannot be derived for the beta distribution but several approximations have been proposed (see Balakrishnan et al. (1995)). We use the method described in Hahn and Shapiro (1994) p. 95.

The ML estimates of the two parameters of the distribution are provided in Table 4. The last two columns report the mean and variance calculated from these estimates. They should be contrasted to the empirical moments available in Table 2. Empirical means seem to be conservative estimates of average recovery rates induced by ML estimation but the empirical standard deviations underestimate the values obtained from ML estimates. Overall it may be observed that even the use of ML estimates cannot dramatically improve the fit when compared to the nonparametric estimates of the previous section.

Table 4: Maximum likelihood estimates of recovery function parameters.

Seniority	$\widehat{\alpha}$	$\widehat{\beta}$	Mean recovery (%)	Standard deviation (%)
Senior secured	1.5529	1.1054	58.42	25.77
Senior unsecured	1.2974	1.3674	48.69	26.11
Subordinated	0.8902	1.3586	39.58	27.13
Junior subordinated	1.1692	1.9157	37.90	24.00
Total	1.0880	1.3318	44.96	26.90

In order to provide a formal judgement about the inadequacy of the beta distribution to describe the stochastic behaviour of the recovery rates, we provide results of goodness-of-fit tests in Table 5. To construct this table we employ a standard parametric bootstrap procedure to approximate the finite sample distribution of a goodness-of-fit test statistic. We use a test statistic based on the integrated squared difference between the kernel density estimator of the unknown density function and the estimator of the parametric density function, here the beta density, to be tested under the null hypothesis of correct specification (Bickel and Rosenblatt (1973), Fan (1994)). If the difference between the nonparametric and parametric estimates is statistically too large, we tend to reject the null hypothesis of correct modelling of recovery rates by a beta distribution. Specifically, the parametric bootstrap testing procedure consists of the following steps (see e.g. Fan (1995)):

1. Estimate the parameter $\theta = (\alpha, \beta)'$ of the beta distribution on the observed recovery rates $\{X_i\}_{i=1}^n$.
2. Compute $\hat{I} = \int (\hat{f} - f_{\hat{\theta}})^2$ where $\hat{f}$, resp. $f_{\hat{\theta}}$, corresponds to the nonparametric, resp. parametric, estimate on $\{X_i\}_{i=1}^n$.
3. Draw a new sample $\{X_i^*\}_{i=1}^n$, called a bootstrap sample, from the estimated beta density $f_{\hat{\theta}}$.
4. Use the bootstrap sample to compute $\hat{I}^* = \int (\hat{f}^* - f_{\hat{\theta}^*})^2$ where $\hat{f}^*$, resp. $f_{\hat{\theta}^*}$, corresponds to the nonparametric, resp. parametric, estimate on $\{X_i^*\}_{i=1}^n$.
5. Repeat Steps 3 and 4 for a number of times, say S times, and obtain the empirical distibution of $\hat{I}_1^*, ..., \hat{I}_S^*$, called the bootstrap distribution.
6. Reject the null hypothesis of correct specification at significance level 5% if $\hat{I}$ is larger than the 95% percentile of the bootstrap distribution.

This type of bootstrap procedures is known to work extremely well when the number of available data is limited.

Results based on 1000 bootstrap samples are shown in Table 5. For many industries, the difference between beta-fitted and kernel estimated densities are

statistically significant at the 5% and even the 1% level. We can thus reject the correct specification of a beta distribution for recovery rates in those cases.

Table 5: Testing the statistical significance of density differences

Industry	90th percentile	95th percentile	99th percentile	$\widehat{I}$
Utility	0.0571	0.0681	0.0846	0.0225
Insurance	0.0809	0.0992	0.1444	0.1072**
Telecom	1.8447	2.4412	3.7342	0.6059
Transport	0.0744	0.0899	0.1369	0.1088**
Financial	0.1423	0.168	0.2596	0.2700***
Chemicals	0.0928	0.1107	0.1693	0.0335
High tech	0.1333	0.161	0.2176	0.0821
Automotive	0.0444	0.0545	0.0699	0.0570**
Building	0.0758	0.0932	0.1419	0.0937**
Consumer	0.0289	0.035	0.0506	0.0502**
Leisure	0.0385	0.0489	0.0684	0.1125***
Energy	0.0489	0.0577	0.0850	0.1371***

* Densities are significantly different at 10%, ** at 5%, *** at 1%.

5.2 Estimating portfolio losses

In this section, we want to assess the impact of the assumption of beta distributed recoveries on the credit Value at Risk (credit VaR) of a well diversified portfolio. The credit VaR at a given loss probability level α and at a chosen horizon (usually one year for credit risk) is the level of losses on the portfolio that will only be exceeded $(1-\alpha)\%$ of the times on average over that horizon. We calculate credit VaR on a simulated portfolio of non investment grade (NIG) bonds in the same industry. We only consider losses due to default and ignore the impact of other transitions or changes in spreads. The number of exposures in the portfolio is assumed to be 1000, each with a unit dollar invested. We assume that all obligors have a marginal default probability of 5.3% and pairwise default correlation of 2.8% which corresponds to the historical averages in non investment grade (NIG) issuers (see De Servigny and Renault (2003a)).

We simulate 10,000 realisations of this portfolio with recovery rates drawn either from the beta distribution calibrated on the empirical mean and variance[2] or from the empirical distribution of the recovery rates [3]. In our example, we choose the Consumer sector. Default indicator functions $X_i,\ i = 1, ..., 1000$ are generated by drawing correlated standard normal random variables Y_i and defining $X_i = I\{Y_i < z\}$, where $I\{.\}$ is an indicator function, $z = N^{-1}(5.3\%)$ is the

[2] We also performed the same calculations with maximum likelihood calibration of the beta distribution. The results (not reported here) were very similar.

[3] We may also use a smoothed distribution based on the nonparametric estimation of Section 4.3. Both procedures deliver similar estimates of credit VaR (they are in fact asymptotically equivalent) but a direct use of the historical dataset is computationally more efficient.

default threshold, calculated by applying the inverse of the cumulative normal distribution function to the probability of default. This standard procedure is explained in more details in Arvanitis et al. (1998).

INSERT FIGURE 3 HERE

The results of the simulations are presented in Figure 3. The figure is a plot of credit VaR at various loss probability levels from 95% to 99.9%, which corresponds to the usual interval considered by banks. The right column corresponds to the credit loss when recovery rates are assumed constant and equal to the mean (36.8%). The middle column is obtained when recoveries are drawn from the beta distribution. It is substantially higher than the level of loss derived from constant recoveries. This highlights the crucial importance of incorporating stochastic recovery rates in credit portfolio models. The left column is computed by drawing recoveries from the empirical distribution (Figure 2). Although the spread between the two columns on the right hand side is smaller than that observable between the columns on the left, use of empirical recoveries can generate credit VaR that are significantly higher (or lower in some industries) than those computed from the simple beta density. In our example, at the 99.9% loss probability level, the "empirical credit VaR" is over 20% higher than the "beta credit VaR". At loss probability levels below 99%, the assumption of beta-distributed recoveries appears to be benign.

As in Section 5.1 we may rely on a standard parametric bootstrap procedure to check whether the difference between the credit VaR estimated under a beta distribution assumption and the credit VaR estimated without this assumption is statistically significant or not. The parametric bootstrap testing procedure becomes in this case:

1. Estimate the parameter $\theta = (\alpha, \beta)'$ of the beta distribution on the observed recovery rates $\{X_i\}_{i=1}^n$.
2. Compute $\hat{T} = \hat{Q} - Q_{\hat{\theta}}$ where $\hat{Q}$, resp. $Q_{\hat{\theta}}$, corresponds to the credit VaR estimated without, resp. with, a parametric assumption on $\{X_i\}_{i=1}^n$.
3. Draw a new sample $\{X_i^*\}_{i=1}^n$, called a bootstrap sample, from the estimated beta density $f_{\hat{\theta}}$.
4. Use the bootstrap sample to compute $\hat{T}^* = \hat{Q}^* - Q_{\hat{\theta}}^*$ where $\hat{Q}^*$, resp. $Q_{\hat{\theta}}^*$, corresponds to the credit VaR estimated without, resp. with, a parametric assumption on $\{X_i^*\}_{i=1}^n$.
5. Repeat Steps 3 and 4 for a number of times, say S times, and obtain the empirical distibution of $\hat{T}_1^*, ..., \hat{T}_S^*$, called the bootstrap distribution.
6. Reject the null hypothesis of no difference between the two credit VaR at significance level 5% if $\hat{T}$ is larger than the 97.5% percentile of the bootstrap distribution or smaller than the 2.5% percentile of the bootstrap distribution.

Table 6 reports the results of the bootstrap test ($S = 1000$). We observe that VaR measures using a beta distribution at high loss probability levels (99.5% and 99.9%) are statistically significantly different from VaRs calculated using empirical recovery rates. At lower loss probability levels (95% and 99%), these differences are no longer significant since the test statistic is inside the bounds of the bootstrap confidence interval.

Table 6: Testing the statistical significance of credit VaR differences

VaR level	90% interval	95% interval	99% interval	$\widehat{T}$
VaR(95%)	[-0.0006,0.0009]	[-0.0008,0.0010]	[-0.0012,0.0013]	0.0005
VaR(99%)	[-0.0016,0.0020]	[-0.0019,0.0024]	[-0,0025,0.0034]	0.0019
VaR(99.5%)	[-0.0025,0.0032]	[-0.0030,0.0037]	[-0.0042,0.0049]	0.0041**
VaR(99.9%)	[-0.0058,0.0066]	[-0.0070,0.0080]	[-0.0094,0.0099]	0.0403***

* Densities are significantly different at 10%, ** at 5%, *** at 1%.

6 Concluding remarks

In this paper, we have examined a nonparametric estimation of the recovery or loss distribution of defaulted bonds. The estimation procedure relies on an asymmetric kernel approach. We have discussed its properties and compare its performance on simulated data. We have then applied the beta kernel density estimator to a large sample of bond losses recorded by Standard & Poor's. We have found that bond seniority and obligor industry are crucially important for the probabilistic behaviour of recovery rates. In addition our nonparametric density estimates show that recovery rates are far from being beta distributed. We have also found that the inappropriate use of such a parametric assumption for the recovery rate distribution can lead to substantial underestimation of credit VaR. The usual market practice of approximating the recovery function through a beta distribution calibrated on the empirical mean and variance should therefore be considered with caution at high loss probability levels for the credit VaR.

REFERENCES :

- Altman E., A. Resti and A. Sironi (2001) "Analyzing and explaining default recovery rates", Report for the ISDA.
- Altman E. and V. Kishore (1996), "Almost everything you wanted to know about recoveries on defaulted bonds", *Financial Analysts Journal*, Nov.-Dec., 57-64.
- Arvanitis A., C. Browne, J. Gregory and R. Martin (1998), "A credit risk toolbox", *Risk*, December, 50-55.
- Asarnow E. and D. Edwards (1995), "Measuring loss on defaulted bank loans: A 24-year study", *Journal of Commercial Lending*, 77, 11-23.
- Bahar R. and L. Brand (2000), "Recoveries on defaulted bonds tied to seniority rankings", Special Report, Standard & Poor's Risk Solutions.
- Balakrishnan N., N. Johnson and S. Kotz (1995), *Continuous univariate distributions*, John Wiley and Sons, New York.
- Basle Committee on Banking Supervision (2003), "The new Basel capital accord", Consultative document, Bank for International Settlements.
- Bickel P. and M. Rosenblatt (1973), "On some global measures of the deviations of density function estimates", *Annals of Statistics*, 1, 1071-1095.
- Bouezmarni T. and J-M. Rolin (2001), "Consistency of beta kernel density function estimator", forthcoming in *Canadian Journal of Statistics*.
- Bouezmarni T. and O. Scaillet (2003), "Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data", UCL Institut de Statistique DP 0306.
- Chen S. (1999), "beta kernel estimators for density functions", *Computational Statistics & Data Analysis*, **31**, 131-145.
- Crouhy M., D. Galai and R. Mark (2001), *Risk Management*, McGraw-Hill.
- De Servigny A. and O. Renault (2003a), "Correlation evidence", *Risk*, July, 90-94.
- De Servigny A. and O. Renault (2003b), *Measuring and Managing Credit Risk*, forthcoming, McGraw-Hill.
- Devroye L. and L. Gyoerfy (1985), *Nonparametric Density Estimation: The L1 View*, Wiley, New York.

- Duffie D. and K. Singleton (2003), *Credit Risk. Pricing, Measurement and Management*, Princeton University Press, Princeton.
- Fan Y. (1994), "Testing the goodness of fit of a parametric density function by kernel method", *Econometric Theory*, 10, 316-356.
- Fan Y. (1995), "Bootstrapping a consistent nonparametric goodness-of-fit test", *Econometric Reviews*, 14, 367-382.
- Friedman C. and S. Sandow (2003), "Ultimate recoveries", *Risk*, August, 69-73.
- Hahn G. and S. Shapiro (1994), *Statistical Models in Engineering*, John Wiley and Sons, New York.
- Hardle, W. and O. Linton (1994), "Applied nonparametric methods" in R. Engle and D. McFadden eds, *Handbook of Econometrics*, 4, 2295-2339.
- Hu Y-T. and W. Perraudin (2002), "The dependence of recovery rates and defaults", Working Paper, Birkbeck College.
- JP Morgan (1997), CreditMetrics technical document.
- Keisman D. and K. Van de Castle (1999), "Recovering your money: insights into losses from defaults", *Standard & Poor's CreditWeek*, June, 16-34.
- Pagan A. and A. Ullah (1999), *Nonparametric Econometrics*, Cambridge University Press, Cambridge.
- Scaillet O. (2001), "Density estimation using inverse and reciprocal inverse Gaussian kernels", forthcoming in the *Journal of Nonparametric Statistics*.
- Schuermann T. (2003), "What do we know about loss-given-default?", Working Paper, Federal Reserve Bank of New York.
- Scott O. (1979), "On optimal and data-based histograms", *Biometrika*, 66, 605-610.
- Silverman B. (1986), *Density Estimation*, Chapman & Hall, London.
- Wand M. and M. Jones (1995), *Kernel Smoothing*, Chapman & Hall, London.

Figure 1: Densities used in simulation experiment

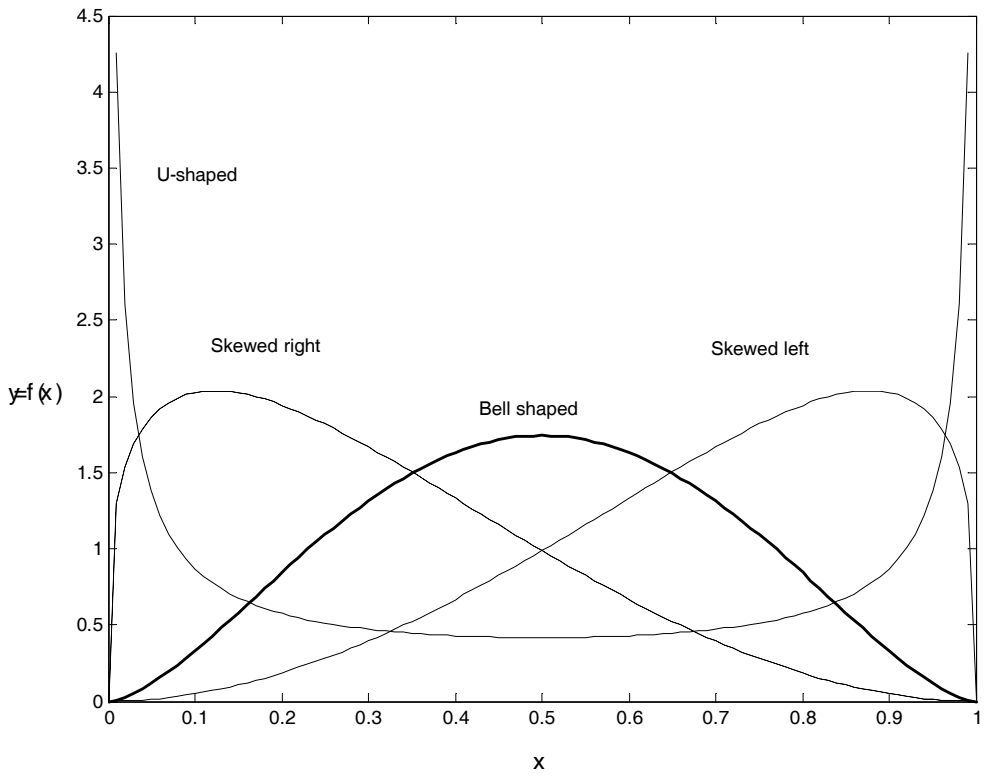


Figure 2(a): Nonparametric and beta fitted recovery distributions per seniority

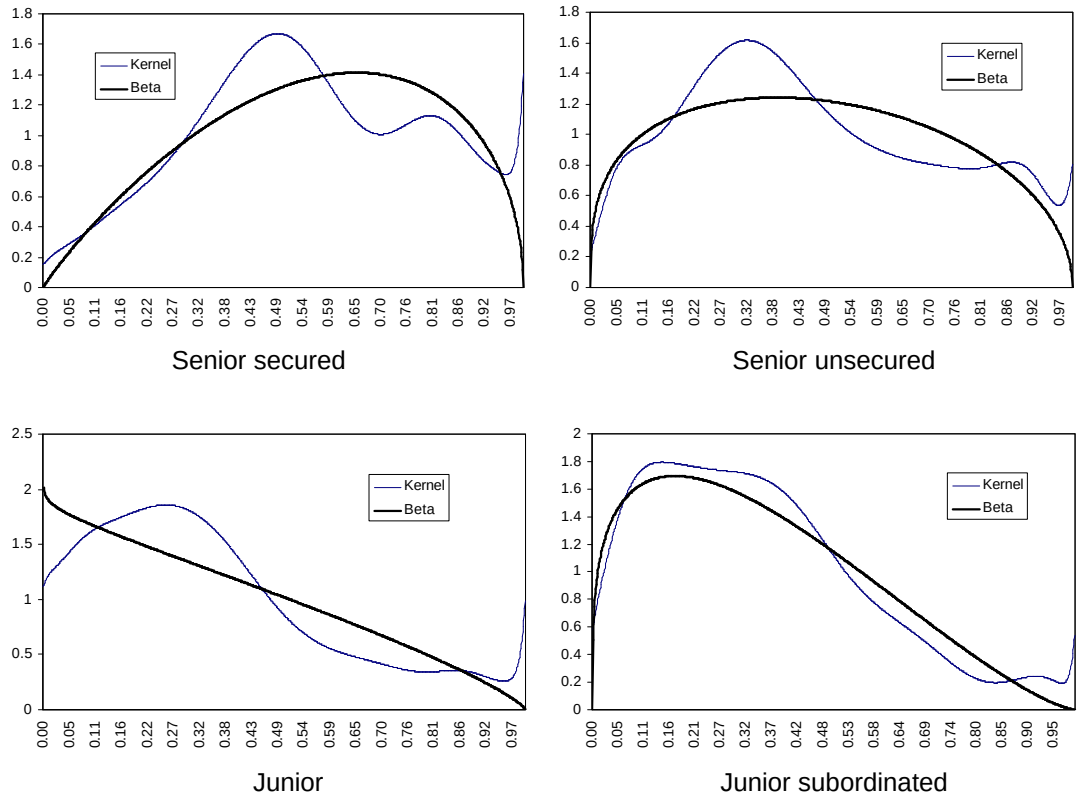
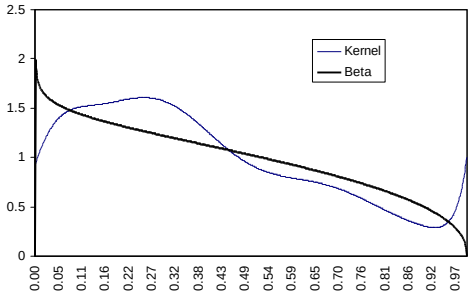
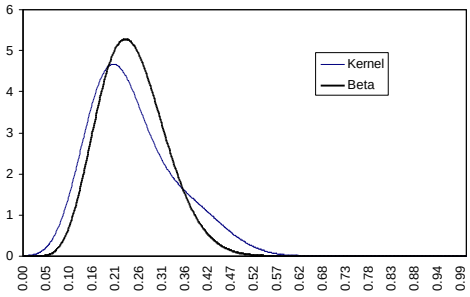


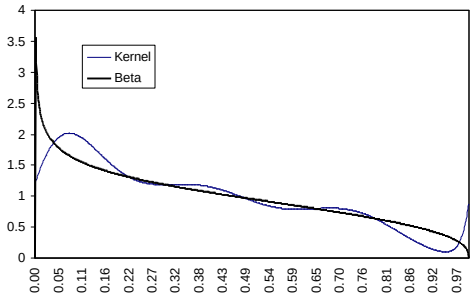
Figure 2(b): Nonparametric and beta fitted recovery distributions per industry



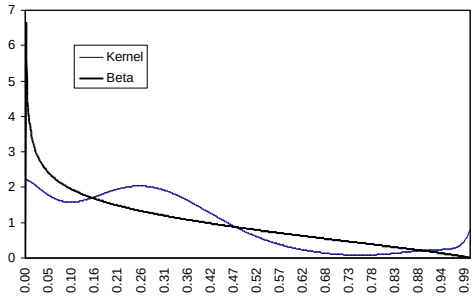
Industry 1



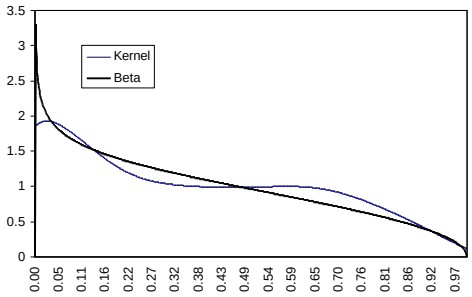
Industry 2



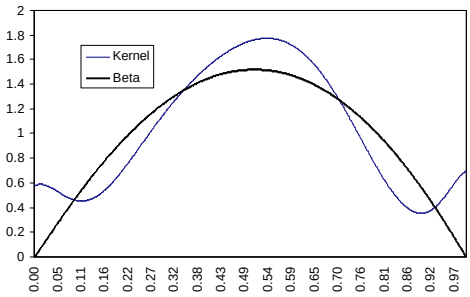
Industry 3



Industry 4



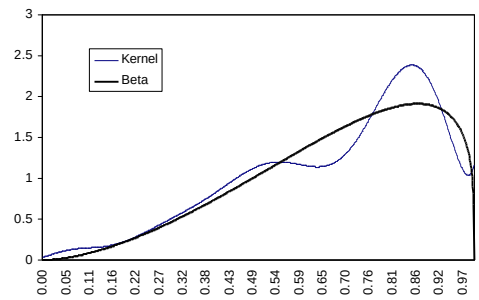
Industry 5



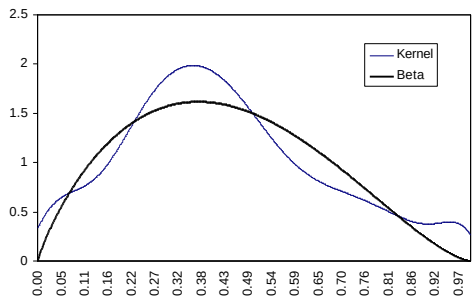
Industry 6

Industry 1: Insurance / real estate
Industry 2: Telecoms
Industry 3: Transportation
Industry 4: Financial
Industry 5: Healthcare / chemical
Industry 6: High technology

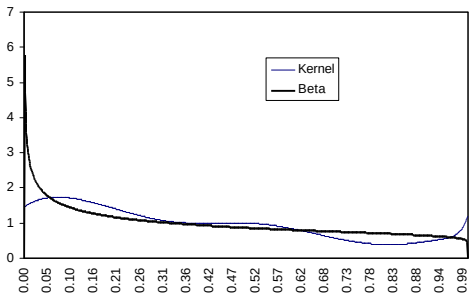
Figure 2(c): Nonparametric and beta fitted recovery distributions per industry



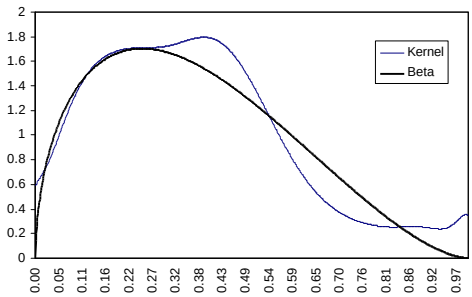
Industry 7



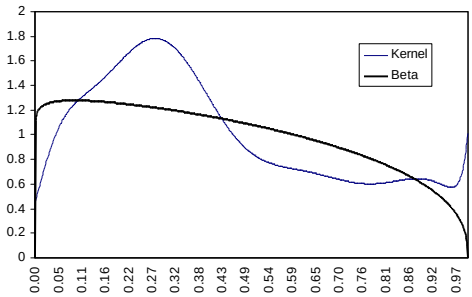
Industry 8



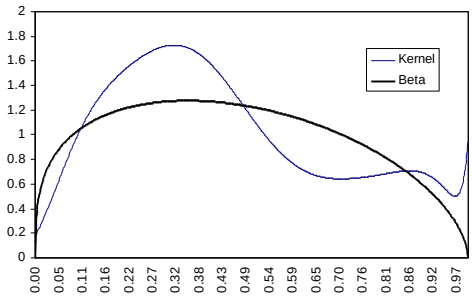
Industry 9



Industry 10



Industry 11



Industry 12

Industry 7: Utility
Industry 8: Aerospace / automotive
Industry 9: Forest / building
Industry 10: Consumer / services
Industry 11: Leisure time / media
Industry 12: Energy / natural resources

Figure 3: Credit VaR at various loss probability levels