QUANTITATIVE FINANCE
RESEARCH CENTRE

UNIVERSITY OF
TECHNOLOGY SYDNEY

QUANTITATIVE FINANCE
RESEARCH CENTRE

UTS

THINK.CHANGE.DO

QUANTITATIVE FINANCE RESEARCH CENTRE

Research Paper 181 **August 2006**

Analytic Models of the ROC Curve: Applications to Credit Rating Model Validation

Stephen Satchell and Wei Xia

ISSN 1441-8010 www.qfrc.uts.edu.au

Analytic Models of the ROC Curve:
Applications to Credit Rating Model Validation

Steve Satchell

Faculty of Economics and Politics

University of Cambridge

Email: ses11@econ.cam.ac.uk

Wei Xia †

Birkbeck College

University of London

E-mail: wxia@econ.bbk.ac.uk

August,2006.

Abstract: In this paper, the authors use the concept of the population ROC curve to build analytic models of ROC curves. Information about the population properties can be used to gain greater accuracy of estimation relative to the non-parametric methods currently in vogue. If used properly this is particularly helpful in some situations where the number of sick loans is rather small; a situation frequently met in periods of benign macro-economic background.

Keywords: Validation, Credit Analysis, Rating Models, ROC, Basel II

Acknowledgement: † Wei Xia would like to thank Birkbeck College for the generous funding support and Ron Smith for helpful comments.

1 Introduction

The Internal Rating-Based (IRB) credit risk modelling approach is now allowed for qualified banks in their risk modelling and economical capital calculation, after the Basel Committee on Banking Supervision (BCBS) published the International Convergence of Capital Measurement and Capital Standards: A Revised Framework, more commonly known as Basel II, in June 2004. One of the important components of most IRB risk models is the rating system used for transforming and assigning the Probability of Default (PD) for each obligor in the credit portfolio. Banks and public rating agencies have developed a variety of rating methodologies in the last three decades. Therefore, questions arise as to which of these methodologies deliver acceptable discriminatory power between the defaulting and non-defaulting obligor ex ante, and which methodologies are to be preferred for different obligor sub-groups. It has become increasingly important for both the regulator and the banks to quantify and judge the quality of rating systems.

This concern is reflected and stressed in the recent BCBS working paper No.14 (2005). This summarizes a number of statistical methodologies for the assessment of discriminatory power which have been suggested in the literature. For example, Cumulative Accuracy Profile (CAP), Receiver Operating Characteristic (ROC), Bayesian error rate, Conditional Information Entropy Ratio (CIER), Kendall's τ and Somers' D, Brier score inter alia. Among those methodologies, the most popular ones are CAP and its summary index, the Accuracy Ratio (AR) as well as ROC and its summary index; this index is called the Area Under the ROC (AUROC). It is worth noting that, unlike some other measures that do not take sample size into account and are therefore substantially affected by statistical errors, the CAP and the ROC measures explicitly account for the size of the default sample and thus can be used for direct rating model comparison.

A detailed explanation of the CAP is presented in Sobehart, Keenan and Stein (2000), and Sobehart and Keenan (2004). ROC has been long used in medicine, psychology and signal detection theory. There is a large body of literature that analyses the properties of the ROC curve. Bamber (1975) shows the AUROC is related to the

Mann-Whitney Statistic and several different methods for constructing the confidence intervals are also discussed. An overview of possible applications of the ROC curves is given by Swets (1988). Sobehart and Keenan (2001) introduced the ROC concept to internal rating model validation. They focus on the calculation and interpretation of the ROC measure. Engelmann, Hayden and Tasche (2003) showed that AR is a linear transformation of AUROC and they complement Sobehart and Keenan (2001) with more statistical analysis of the ROC. However, previous works that we are familiar with, have been done using a finite sample of empirical data or simulated data. None of the above have analyzed the analytic properties of the ROC Curve and the ROC measure under parametric assumptions concerning the distribution of the rating scores.

In this paper, we further explore the statistical properties of the ROC Curve and its summary indices, especially under a number of rating score distribution assumptions. We focus on the analytical properties of ROC Curve only since the CAP measure is just a linear transformation of the ROC measure.

In section 2, in order to keep this paper self-contained, we first briefly introduce the credit rating model validation background and explain the concepts and definitions of ROC and CAP. A general equation for the ROC Curve is derived. By assuming that there exists probability density functions of the two variables that construct the ROC Curve, an unrestrictive assumption, we show that there is a link between the first derivative of the curve and the likelihood ratio of the two variables, a result derived by different methods in Bamber(1975).

In section 3, by further assuming certain statistical distributions for the credit rating scores, we derive analytic solutions for the ROC Curve and its summary indices respectively. In particular, when the underlying distributions are both Negative Exponential distributions, we have a closed form solution.

In section 4, we apply the results derived in section 3 to simulated data. Performance evaluation reports are presented. Section 5 concludes:

We find that estimation results by our analytic approach are as good as and, in many cases, better than the non-parametric AUROC ones. Although the accuracy of the approach in this paper is limited by the continuous rating score assumption and also affected by the accuracy of estimation of the distribution parameters on rating score samples in some cases, it offers direct insight into more complex situations and a better tool in credit rating model selection procedure since the analytic solution can be used as object function.

2 Theoretical Implication and Applications:

In this section, we first briefly review the credit rating system methodology, in particular, the CAP and the ROC measures. The content presented in this part is very similar to Engelmann, Hayden and Tasche (2003) and BCBS working paper No.14 (2005). Then we introduce the Ordinary Dominance Graph (ODG) where ROC is a special case of ODG and some interesting theoretical implications of the ROC curve will be given.

2.1 The validation of credit rating system.

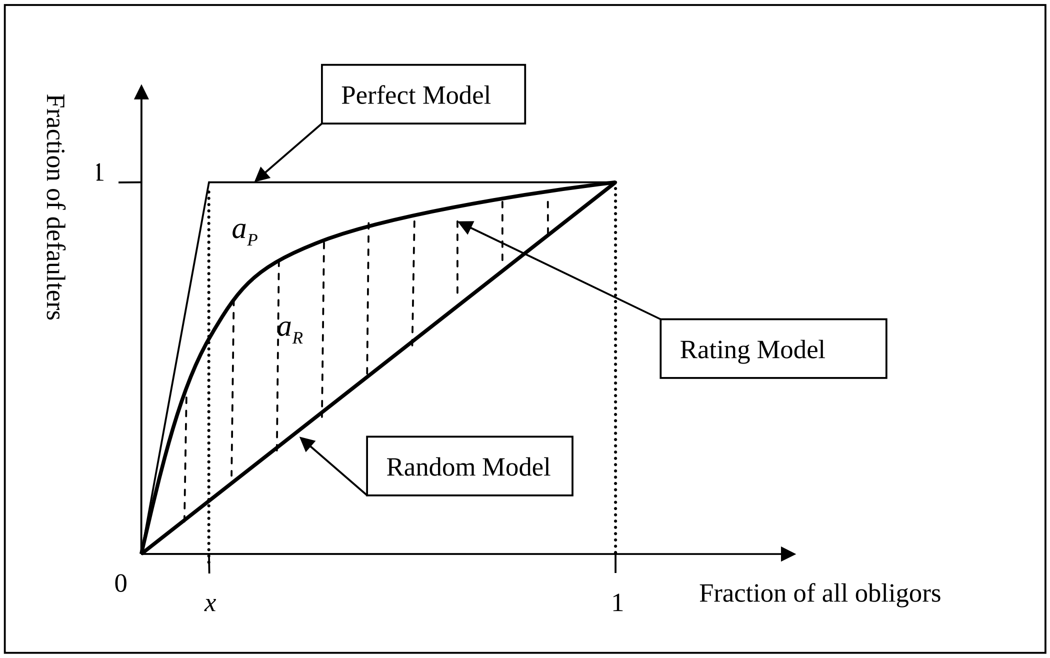
The statistical analysis of rating models is based on the assumption that for a predefined time horizon there are two groups of a bank's obligor: obligors that will be in default, called defaulters, and obligors that will not be in default, called non-defaulters. It is not observable in advance whether an obligor is a defaulter or a non-defaulter in the next time horizon. Banks have loan books or credit portfolio, they have to assess an obligor's future status based on a set of his or her present observable characteristics. Rating systems may be regarded as classification tools to provide signals and indications of the obligor's possible future status. A rating score is returned for each obligor based on a rating model, usually an Expert Judgment Model. The main principal of rating systems is that “the better a grade, the smaller the proportion of defaulters and the greater the proportion of non-defaulters that are assigned this grade”. Some examples are the famous Altman’s Z score, or some score from a Logit model or from other approach.

Therefore, the quality of a rating system is determined by its discriminatory power between non-defaulting obligors and defaulters ex ante for a specific time horizon, usually a year. The CAP measure and ROC provide statistical measures to assess the discriminatory power of various rating models based on historical data.

2.2 Cumulative Accuracy Profile and Accuracy Ratio

Consider an arbitrary rating model that produces a rating score. A high rating score is usually an indicator of a low default probability. To obtain the CAP curve, all debtors are first ordered by their respective scores, from riskiest to safest, i.e. from the debtor with the lowest score to the debtor with the highest score. For a given fraction x of the total number of debtors the CAP curve is constructed by calculating the percentage $d(x)$ of the defaulters whose rating scores are equal to or lower than the maximum score of fraction x. This is done for x ranging from 0% to 100%. Figure 1 illustrates CAP curves.

Figure 1. Cumulative Accuracy Profile



Apparently a perfect rating model will assign the lowest scores to the defaulters. In this case the CAP is increasing linearly and than staying at one. For a random model without any discriminative power, the fraction *x* of all debtors with the lowest rating scores will contain *x* percent of all defaulters. Real rating system lies somewhere in between these two extremes. The quality of a rating system is measured by the Accuracy Ratio (AR). It is defined as the ratio of the area a_R between the CAP of the rating model being validated and the CAP of the random model, and the area a_P between the CAP of the perfect rating model and the CAP of the random model.

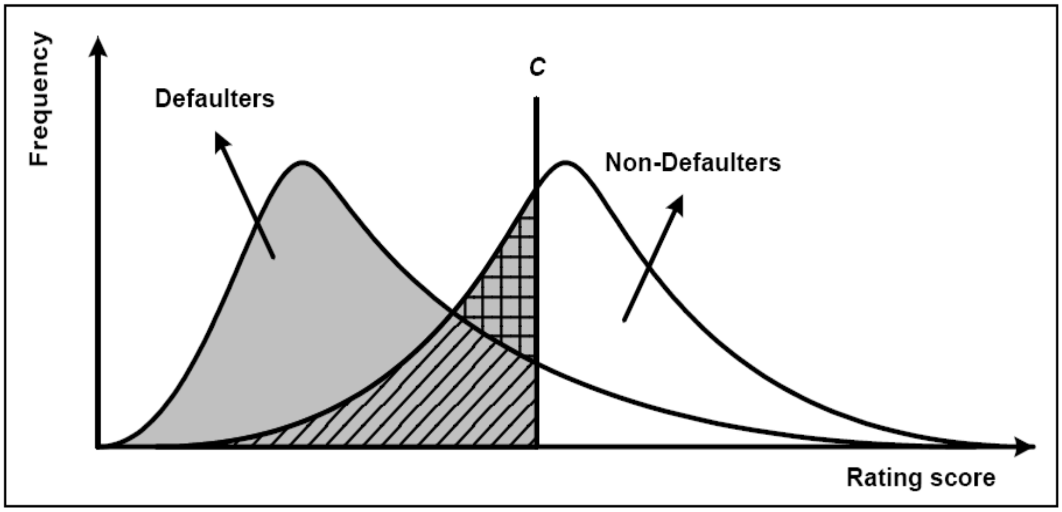
$$AR = \frac{a_R}{a_P}$$

It is easy to see that for real rating models the AR range from zero to one and the rating model is the better if AR is closer to one.

2.3 Receiver Operating Characteristic and the Area Under the ROC curve

The construction of a ROC curve is illustrated in Figure 2 which shows possible distributions of rating scores for defaulting and non-defaulting debtors. For a perfect rating model the left distribution and the right distribution in Figure 2 would be separate. For real rating systems, perfect discrimination in general is not possible. Both distributions will overlap as illustrated in Figure 2.

Figure 2: Distribution of rating scores for defaulting and non-defaulting debtors



Assume someone has to find out from the rating scores which debtors will survive during the next period and which debtors will default. One possibility for the decision-maker would be to introduce a cut-off value *C* as in Figure 2, and to classify each debtor with a rating score lower than *C* as a potential defaulter and each debtor with a rating score higher than *C* as a non-defaulter. Then four decision results would be possible. If the rating score is below the cut-off value *C* and the debtor defaults subsequently, the decision was correct. Otherwise the decision-maker wrongly classified a non-defaulter as a defaulter. If the rating score is above the cut-off value and the debtor does not default, the classification was correct. Otherwise a defaulter was incorrectly assigned to the non-defaulters' group.

Then one can define a hit rate *HR(C)* as:

$$HR(C) = \frac{H(C)}{N_D}$$

where *H(C)* is the number of defaulters predicted correctly with the cut-off value *C*, and N_D is the total number of defaulters in the sample. This means that the hit rate is the fraction of defaulters that was classified correctly for a given cut-off value *C*. The false alarm rate then *FAR(C)* is defined as:

$$FAR(C) = \frac{F(C)}{N_{ND}}$$

where *F(C)* is the number of false alarms, i.e. the number of non-defaulters that were classified incorrectly as defaulters by using the cut-off value *C*. The total number of non-defaulters in the sample is denoted by N_{ND} . In Figure 2, *HR(C)* is the area to the left of the cut-off value *C* under the score distribution of the defaulters (coloured plus hatched area), while *FAR(C)* is the area to the left of *C* under the score distribution of the non-defaulters (chequered plus hatched area).

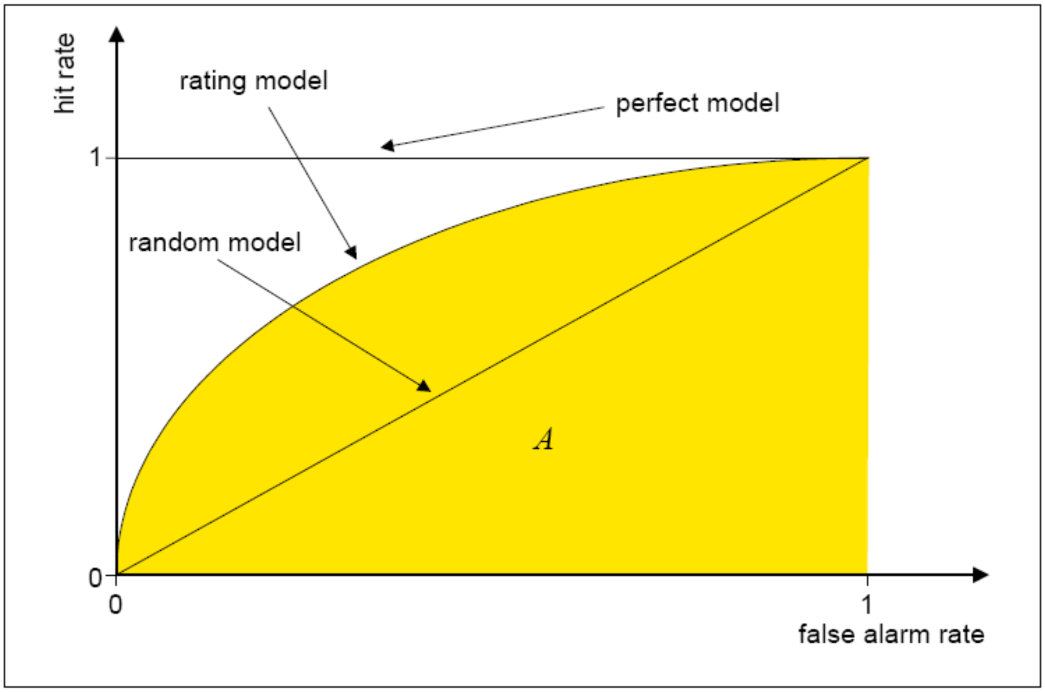
The ROC curve is constructed as follows. For all possible cut-off values *C* that are contained in the range of the rating scores the quantities *HR(C)* and *FAR(C)* are computed. The ROC curve is a plot of *HR(C)* versus *FAR(C)*. This is illustrated in Figure 3.

A rating model's performance is the better the steeper the ROC curve is at the left end and the closer the ROC curve's position is to the point (0,1). Similarly, the model is the better the larger the area under the ROC curve is. This area is called AUROC and is denoted by *A*. By means of a change of variable, it can be calculated as

$$A = \int_0^1 HR(FAR)\,d(FAR)$$

The area *A* is 0.5 for a random model without discriminative power and it is 1.0 for a perfect model. It is between 0.5 and 1.0 for any reasonable rating model in practice.

Figure 3: Receiver Operating Characteristic Curves. (Taken from Tasche. (2003))



It has been shown in Engelmann, Hayden and Tasche (2003) that:

$$AR = \frac{a_R}{a_P} = \frac{N_{ND}\left(A - 0.5\right)}{0.5N_{ND}} = 2\left(A - 0.5\right) = 2A - 1$$

2.4 Some further statistical properties of ROC measures

The ROC stems from the Ordinal Dominance Graph (ODG). Assume we have two sets of continuous random variables, *X* and *Y*. Let C be an arbitrary constant. Define:

$$y = prob\left(Y \le C\right) = F_Y\left(C\right) \text{ and } x = prob\left(X \le C\right) = F_X\left(C\right)$$

, where *x* and *y* lie between [0, 1] and C lies in $\left(-\infty, +\infty\right)$. Then the ODG is simply a plot of *y* against *x*. See Figure 4. There are some straight forward properties of the ODG:

1) The ODG curve is never decreasing as *x* increases *y* cannot decrease.
2) If $Prob(Y \le C) = Prob(X \le C)$, then *x* and *y* are identically distributed, *y* = *x* and the ODG curve is a $45°$ line.

3) If X first order stochastic dominance (FSD) Y, then the ODG curve lies above the $45°$ line and vice versa.

Proof:

If X FSD $Y \Rightarrow F_X(C) \le F_Y(C), \quad \forall C \in \mathbb{R} \Rightarrow x \le y$

$\Rightarrow (x, y)$ lies above the 45° line

If we regard y as score signals of defaulters in the next predefined period and x as those of the non-defaulters, then we expect any sensible rating system produces $Prob(Y \le C) \ge Prob(X \le C)$ for all C. Thus $x \le y$ for all C and the ODG curve is above the $45°$ line. It is referred to as the ROC curve in the risk literature. In terms of section 2.3, y is the *HR(C)* and x is the *FAR(C)*.

By assuming there exists probability density functions (PDF) of F_X and F_Y, i.e. that they are both absolutely continuous, the following lemma can be derived:

Lemma 1:

If $x = F_X(C), C = F_X^{-1}(x)$, then $\frac{\partial C}{\partial x} = \frac{1}{f_X(C)}$, where $f_X(C)$ is the PDF of X..

Proof: $1 = \frac{\partial F_X(C)}{\partial x} = \frac{\partial F_X(C)}{\partial C} . \frac{\partial C}{\partial x} = f_X(C) \frac{\partial C}{\partial x}, \ \therefore \frac{\partial C}{\partial x} = \frac{1}{f_X(C)}$

Lemma 2:

If $y = F_Y(C)$, then $\frac{\partial y}{\partial x} = \frac{f_Y(C)}{f_X(C)}$.

We see from Lemma 2 that the slope of the ODG curve is just the likelihood ratio (LR) of Y and X evaluated at C.

Proof: $\frac{\partial y}{\partial x} = \frac{\partial F_Y(C)}{\partial x} = \frac{\partial F_Y(C)}{\partial C} . \frac{\partial C}{\partial x} = f_Y(C) . \frac{\partial C}{\partial x} = \frac{f_Y(C)}{f_X(C)}$

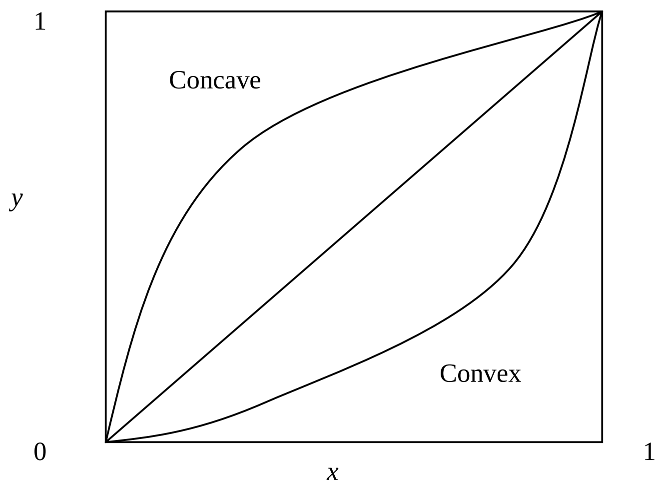
Theorem:

If $\dfrac{f_Y(C)}{f_X(C)}$ is increasing in C, then $\dfrac{\partial y}{\partial x}$ is increasing $\Rightarrow$ the ODG curve is convex.

If $\dfrac{f_Y(C)}{f_X(C)}$ is decreasing in C, then $\dfrac{\partial y}{\partial x}$ is decreasing $\Rightarrow$ the ODG curve is concave.

The later case is the one that we are interested in, as it is the ROC curve. See Figure 4.

Figure 4: Ordinal Dominance Graph



We have assumed that X and Y have PDF's, thus they are continuous random variables. The AUROC can then be expressed as:

A = Prob($y \le x$) = $\int_{-\infty}^{\infty} \text{Prob}\left(Y \le X \mid X = C\right)\text{Prob}\left(X = C\right)dC$

Since X and Y are scores from different obligor groups, they are independent.

We have: $\text{Prob}\left(Y \le X \mid X = C\right) = \text{Prob}\left(Y \le C\right)$.

Since $y = F_Y(C) = \text{Prob}\left(Y \le C\right)$ and $\partial x = f_X(C)\partial C$ $\Rightarrow$

$$A = \int_{-\infty}^{\infty} F_Y(\text{C})\, f_X(C)\, dC = \int_0^1 F_Y\left(F_X^{-1}(x)\right)dx \quad \text{------------------------------ (1)}$$

A modelling exercise may choose a rating model that maximizes the AUROC with respect to the obligor group under review. But how would one estimate Prob($y \leq x$)? Bamber(1975) and Engelmann, Hayden and Tasche (2003) show that given a rating scores sample of the obligors assigned by a rating model, the AUROC could be estimated in an non-parametric way using the Mann-Whitney U statistic.

On the other hand, if we had a parametric distribution model for *X* and *Y*, then we could explicitly derive a formula for the ROC curve and for the AUROC. In the next section, we will review some plausible distributions for *X* and *Y* and derive the close form solutions wherever possible.

3 Choices of Distributions

In this section, we derive analytical formulae for the ROC by assuming that the rating scores produced by rating models follow some plausible distributions. The distributions we presented here are Weibull Distribution (including Exponential Distribution), Logistic Distribution, Normal Distribution and Mixed models for X and Y respectively. In the cases that we have explicit closed forms for the ROC curve, we derive the closed form AUROC as well. The case of mixed distributions for X and Y can be easily extended from the discussion followed.

We use the symbol *M* for the location parameters (sometimes the mean, sometimes the minimum), λ for the scale parameter and α for the shape parameter in the distribution functions where appropriate.

3.1 Weibull Distribution

We first present solutions under a Weibull Distribution assumption of rating scores. The Weibull Distribution is flexible and rich. A three parameter Weibull distribution cumulative probability function (CDF) is:

$$F(z) \equiv P(Z \le z) = 1 - e^{-\left(\frac{z-M}{\lambda}\right)^{\alpha}},$$

where $z>M$, $\alpha>0$ and $\lambda>0$ The inverse CDF of a three parameter Weibull Distribution is:

$$F^{-1}(p) = M + \lambda\left[-\ln(1-p)\right]^{\frac{1}{\alpha}},$$

where $p \in [0,1]$. Assuming *y* is the *HR(C)* and *x* is the *FAR(C)*, the three-parameter Weibull Distribution ROC is derived as:

$$x = F_X(C) = 1 - e^{-\left(\frac{C-M_x}{\lambda_x}\right)^{\alpha_x}} \Rightarrow C = M_x + \lambda_x\left[-\ln(1-x)\right]^{\frac{1}{\alpha_x}}$$

$$y = F_Y(C) = P(Y \le C) = 1 - e^{-\left(\frac{C-M_y}{\lambda_y}\right)^{\alpha_y}}$$

$$\text{ROC}: y = 1 - \exp\left[-\left(\frac{\lambda_x}{\lambda_y}\left[-\ln(1-x)\right]^{\frac{1}{\alpha_x}} + \left(M_x - M_y\right)\right)^{\alpha_y}\right] \text{ --------------------------- (2)}$$

The above ROC formula is very difficult to use to deduce an explicit closed form formula for the AUROC, although a numerical solution exists in this case once all the parameters are estimated. However, the situation becomes much better if we impose a slightly stronger assumption that the shape parameters α of the F_X and F_Y are equal to one. We then have analytical closed form solutions in this case and the Weibull distribution degenerates to a truncated Exponential Distribution Family if M is positive and an extended Exponential if M is negative.

Assume: $\alpha_x = \alpha_y = 1$, we then rewrite the above equation as following:

$$\text{ROC}: y = 1 - e^{\frac{M_y - M_x}{\lambda_y}}(1-x)^{\frac{\lambda_x}{\lambda_y}}$$

Let $K = e^{\frac{M_y - M_x}{\lambda_y}}$, $\theta = \frac{\lambda_x}{\lambda_y}$, then

$$AUROC = \int_0^1\left[1 - K(1-x)^{\theta}\right]dx = 1 - \frac{K}{1+\theta} = 1 - \frac{\lambda_y}{\lambda_x + \lambda_y}e^{\frac{M_y - M_x}{\lambda_y}}, \text{ ----------------------- (3)}$$

We next discuss some of the properties of equation 3.

Property 1.

$M_y - M_x \in (-\infty, 0)$ for any plausible rating system. The smaller is $M_y - M_x$ (or the larger is $\left|M_y - M_x\right|$), the closer is the AUROC to one. Recall that M is the location parameter, in this case the minimum. Therefore, the rating system will receive a higher AUROC if it can better discriminate the defaulter from the non-defaulter by the difference in their minimum values.

It is also interesting to see that if $M_y - M_x \to 0$, $\text{AUROC} \to 1 - \frac{\lambda_y}{\lambda_x + \lambda_y}$. As we illustrated in early section that AUROC of a plausible rating system is above 0.5 if it is not a random selection system. This implies that the value of the scale parameters that the rating scores being assigned have to be such that $0 < \lambda_y \le \lambda_x$ in this case. Note that this condition is implied by both groups being exponential but also by both groups being truncated or extended exponentials with the same minima.

Property 2.

AUROC is monotonically increasing with respect to λ_x, but monotonically decreasing with respect to λ_y.

Proof:

$$\frac{d}{d\lambda_y} AUROC = \frac{K}{\left(\lambda_x + \lambda_y\right)^2 \lambda_y}\left[\left(M_y - M_x\right)\left(\lambda_x + \lambda_y\right) - \lambda_y\left(\lambda_x + 2\lambda_y\right)\right]$$

$$\left.\begin{array}{l}\left[\left(M_y - M_x\right)\left(\lambda_x + \lambda_y\right) - \lambda_y\left(\lambda_x + 2\lambda_y\right)\right] < \left(\lambda_x + \lambda_y\right)\left(M_y - M_x - \lambda_y\right) < 0 \\ \text{Since plausible rating systems, we expect } M_y \le M_x,\ \mathrm{K} \ge 0 \text{ and } \lambda_y > 0\end{array}\right\} \Rightarrow$$

$$\Rightarrow \frac{d}{d\lambda_y} AUROC \le 0$$

3.2 Logistic Distribution

A two parameter Logistic distribution CDF is:

$$F(z) \equiv P(Z \le z) = \frac{e^{\frac{z-M}{\lambda}}}{1+e^{\frac{z-M}{\lambda}}},$$

where $z \in \mathbb{R}$ and $\lambda > 0$. Here M is a mean parameter.

The inverse CDF of a two parameter Logistic Distribution is:

$$F^{-1}(p) = M + \lambda \ln\left(\frac{p}{1-p}\right),$$

where $p \in [0,1]$. Again assuming y is the *HR(C)* and x is the *FAR(C)*, we have:

$$x = F_X(C) = \frac{e^{\frac{C-M_x}{\lambda_x}}}{1+e^{\frac{C-M_x}{\lambda_x}}} \Rightarrow C = M_x + \lambda_x \ln\left(\frac{x}{1-x}\right) \qquad \text{and}$$

$$y = F_Y(C) = F_Y\left(F_X^{-1}(x)\right) = \frac{e^{\frac{C-M_y}{\lambda_y}}}{1+e^{\frac{C-M_y}{\lambda_y}}} = \frac{e^{\frac{M_x-M_y}{\lambda_y}}\left(\frac{x}{1-x}\right)^{\frac{\lambda_x}{\lambda_y}}}{1+e^{\frac{M_x-M_y}{\lambda_y}}\left(\frac{x}{1-x}\right)^{\frac{\lambda_x}{\lambda_y}}}$$

Similar to the Weibull Distribution case, the AUROC with the above ROC specification can always be evaluated numerically. Moreover, by assuming $\lambda_x = \lambda_y = 1$, and in what follows assume that K does not equal 1, $K = e^{M_x - M_y}$, the above ROC equation can be simplified to

$$y = \frac{e^{M_x-M_y}x}{1-x+e^{M_x-M_y}x} = \frac{Kx}{1+(K-1)x}$$

The AUROC can be now derived analytically.

$$AUROC = K\int_0^1 \frac{x}{1+(K-1)x}dx = \frac{K}{K-1}\int_0^1 \frac{(K-1)x}{1+(K-1)x}dx$$

Let $u = 1 + (K-1)x$

$$AUROC = \frac{K}{(K-1)^2}\int_1^K \left(1 - \frac{1}{u}\right) du = \frac{K}{K-1}\left(1 - \frac{\ln K}{K-1}\right) \Rightarrow \lim_{K\to\infty} AUROC = 1$$

3.3 Normal Distribution

A two parameter Normal distribution CDF is:

$$F(z) \equiv P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\lambda\sqrt{2\pi}} e^{\frac{1}{2}\left(\frac{x-M}{\lambda}\right)^2} dx = \Phi\left(\frac{z-M}{\lambda}\right),$$

where $z \in$ and the $\Phi(.)$ is the standard Normal probability distribution function..

The inverse CDF of a two parameter Logistic Distribution is: $F^{-1}(p) = M + \lambda\Phi^{-1}(p)$,

where $p \in [0,1]$. For *y* is the *HR(C)* and *x* is the *FAR(C)*, we have:

$$x = \Phi\left(\frac{C - M_x}{\lambda_x}\right) \Rightarrow C = M_x + \lambda_x \Phi^{-1}(x)$$

$$y = \Phi\left(\frac{C - M_y}{\lambda_y}\right) = \Phi\left(\frac{(M_x - M_y) + \lambda_x \Phi^{-1}(x)}{\lambda_y}\right)$$, which gives the ROC curve.

Therefore, $AUROC = \int_0^1 \Phi\left(\frac{(M_x - M_y) + \lambda_x \Phi^{-1}(x)}{\lambda_y}\right) dx$

This function can be easily evaluated numerically to obtain the AUROC.

Property 1

AUROC increases with $M_x - M_y$ and in particular if $M_x - M_y = 0$, and $\lambda_x = \lambda_y$.,AUROC = 0.5

Proof:

$$AUROC = \int_0^1 \Phi\left(\frac{(M_x - M_y) + \lambda_x \Phi^{-1}(x)}{\lambda_y}\right) dx = \int_0^1 \Phi\left(\frac{\lambda_x}{\lambda_y}\Phi^{-1}(x)\right) dx = \int_0^1 x dx = 0.5$$

The above property is also quite intuitive. If the means and variances of two normal distributed populations equal each other,the distributions are equal and then overall there is no discriminatory power of the models based on this rating mechanism. So the AUROC is 0.5.

3.4 Mixed Models

It is obvious that as long as we have parametric distribution families for the defaulter and non-defaulters, one can always calculate an AUROC for the two score samples from equation (1) in section 2 even with two different parametric distributions for the two populations respectively.

4. Performance Evaluation on the AUROC Estimation with Simulated Data

In this section we carry out a performance evaluation on AUROC estimations by the non-parametric Mann-Whitney Statistic and the analytic approach suggested in this paper respectively with simulated data.

We first assume known parametric distributions for the credit scores of defaulters and non-defaulters. By doing this we would know the value of the theoretical AUROC. After generating simulated sample data from the assumed distributions for defaulter score and non-defaulter's, we estimate the AUROC and its confidence interval (CI) by the above two approaches on the simulated samples. We repeat the simulation and estimation procedure a number of times. We then compare the accuracy of the AUROC estimation and the CI of the two approaches. Finally, we change the parameter values of the assumed distribution and repeat the simulation. We repeat the above procedures to evaluate the performance of the two approaches subject to different theoretical AUROC index values with different sample sizes of defaulters. We choose two-parameter Normal Distributions, one-parameter Exponential Distributions and Weibull Distributions with various shape and scale parameters.

4.1 Performance evaluations under Normal Distribution Assumption

Normal Distributions are assumed as our parametric distributions of the credit scores of the defaulters and the credit scores of the non-defaulters in this performance evaluation. The theoretical value of AUROC for the normal score samples is evaluated numerically①. The non-parametric estimate of AUROC is carried out using the ROC module in SPSS and we use the bootstrap to re-sample 1000 replications to obtain the estimates of the analytic approach which also generates a two-sided 95% CI. The parameters of the parametric distribution are estimated for each replication and substituted back to the analytic AUROC formula. We then define error as the difference between model estimates based on a sample and the theoretical AUROC value, and compare the mean error and mean absolute error for the two approaches. The width of the confidence interval is also compared.

We generate 50 normal samples from six different settings respectively. Settings 1, 2 and 3, consisting of Group 1, target the AUROC at a low value, while settings 4, 5 and 6, Group 2, target the AUROC at a high value. Within each group they vary in defaulter's sample size, which ranges at 20, 100 and 500. Credit rating models can be applied to at least three different types of groups: credit risk with Corporate, counter party default risk in Trading Books, and credit risk in credit card and other loan type Banking Books. The default sample of Corporate is usually small, such as 50 in ten years, especially under a good economic cycle. Meanwhile, the number of defaults in a loan book or a credit card book in a commercial bank's banking book can be fairly large, possibly in excess of several hundreds. The reason for selecting different defaulter sample sizes in the test is to assess for which type of problem the analytic approach outperforms. We define a performance statistic as follows:

(Ratio to N) = Difference / (Non-Parametric Estimate).

Normal Setting 1-3:

Normal Distributions			Setting 1	Setting 2	Setting 3
Sample	Mean	Standard	# of	# of	# of

		deviation	Observation	Observation	Observation
X	2	2	1000	1000	1000
Y	1	3	20	100	500

Theoretical AUROC= 0.609239

Report on estimation error with 50 simulated normal samples & 1000 replication bootstrap

Setting 1

Approach	Non-parametric	Analytic	Difference	Ratio to N
Mean Error	0.007482	0.007466	0.000016	
Mean ABS Error	0.050164	0.048408	0.001756	3.50%
Mean CI Width	0.289225	0.276522	0.012703	4.39%

Setting 2

Approach	Non-parametric	Analytic	Difference	Ratio to N
Mean Error	0.000404	-0.081959	0.082363	
Mean ABS Error	0.025946	0.024957	0.000989	3.81%
Mean CI Width	0.136364	0.130728	0.005636	4.13%

Setting 3

Approach	Non-parametric	Analytic	Difference	Ratio to N
Mean Error	0.002643	0.002752	-0.000109	
Mean ABS Error	0.014172	0.014809	-0.000636	-4.49%
Mean CI Width	0.064965	0.062608	0.002357	3.63%

Normal Setting 4-6:

Normal Distributions			Setting 4	Setting 5	Setting 6
Sample	Mean	Standard deviation	# of Observation	# of Observation	# of Observation
X	2	0.5	1000	1000	1000
Y	1	1	20	100	500

Theoretical AUROC= 0.814448

Report on estimation error with 50 simulated normal samples & 1000 replication bootstrap

Setting 4

Approach	Non-parametric	Analytic	Difference	Ratio to N
Mean Error	0.009138	0.006718	0.002421	
Mean ABS Error	0.046588	0.045725	0.000863	1.85%
Mean CI Width	0.232187	0.215922	0.016265	7.01%

Setting 5

Approach	Non-parametric	Analytic	Difference	Ratio to N
Mean Error	0.001187	0.000510	0.000678	
Mean ABS Error	0.025564	0.024311	0.001253	4.90%
Mean CI Width	0.112444	0.107148	0.005296	4.71%

Setting 6

Approach	Non-parametric	Analytic	Difference	Ratio to N
Mean Error	0.001470	0.001303	0.000167	
Mean ABS Error	0.012239	0.011061	0.001178	9.62%
Mean CI Width	0.052653	0.049464	0.003189	6.06%

In tables 1-6, all mean confidence interval widths show that the estimates of the analytic approach are better then the non-parametric estimates. As for the mean error and the mean absolute error, analytic estimates outperform the non-parametric estimates in tables 1, 2 and 4-6. (Ration to N) shows the percentage of the difference out of the non-parametric approach estimate. The larger the (Ration to N), the more the analytic approach outperform the non-parametric approach.

4.2Performance evaluations under Exponential Distribution Assumption

Exponential Distributions are assumed as our parametric distribution of the credit scores of the defaulters and the credit scores of the non-defaulters in this performance evaluation. The theoretical value of AUROC for the Exponential score samples is evaluated analytically by the closed form formula (3) in section 3.1. The performance evaluation setting is very similar to the Normal Distribution one. There are 6 settings across different AUROC value and defaulter sample size as well.

Exponential Setting 1-3:

Normal Distributions		Setting 1	Setting 2	Setting 3
Sample	Scale Parameter (Lamda)	# of Observation	# of Observation	# of Observation
X	3	1000	1000	1000
Y	1.5	20	100	500

Theoretical AUROC=0.666667

Report on estimation error with 50 simulated normal samples & 1000 replication bootstrap

Setting 1

Approach	Non-parametric	Analytic	Difference	Ratio to N
Mean Error	-0.008179	-0.007035	-0.001144	
Mean ABS Error	0.040993	0.040056	0.000938	2.29%
Mean CI Width	0.209540	0.189586	0.019954	9.52%

Setting 2

Approach	Non-parametric	Analytic	Difference	Ratio to N
Mean Error	-0.000987	0.000034	-0.001021	
Mean ABS Error	0.025320	0.021922	0.003398	13.42%
Mean CI Width	0.099043	0.088280	0.010763	10.87%

Setting 3

Approach	Non-parametric	Analytic	Difference	Ratio to N
Mean Error	-0.002926	-0.003401	0.000475	
Mean ABS Error	0.011471	0.011015	0.000456	3.98%
Mean CI Width	0.055636	0.047672	0.007964	14.31%

Exponential Setting 4-6:

Normal Distributions		Setting 1	Setting 2	Setting 3
Sample	Scale Parameter (Lamda)	# of Observation	# of Observation	# of Observation
X	4	1000	1000	1000
Y	1	20	100	500

Theoretical AUROC=0.800000

Report on estimation error with 50 simulated normal samples & 1000 replication bootstrap

Setting 4

Approach	Non-parametric	Analytic	Difference	Ratio to N
Mean Error	-0.008576	-0.006721	-0.001855	
Mean ABS Error	0.033790	0.031174	0.002616	7.74%
Mean CI Width	0.145500	0.132758	0.012742	8.76%

Setting 5

Approach	Non-parametric	Analytic	Difference	Ratio to N
Mean Error	0.002783	0.003403	-0.000621	
Mean ABS Error	0.015655	0.014320	0.001335	8.53%
Mean CI Width	0.071140	0.064132	0.007008	9.85%

Setting 6

Approach	Non-parametric	Analytic	Difference	Ratio to N
Mean Error	0.000118	0.000521	-0.000403	
Mean ABS Error	0.007710	0.007495	0.000215	2.79%
Mean CI Width	0.043742	0.034280	0.009462	21.63%

In table 1-6, all the mean absolute error and the mean confidence interval widths show that the estimates of the analytic approach are better then the non-parametric estimates. (Ration to N) shows that the non-parametric approach estimates provide a significantly better confidence interval than the non-parametric estimates.

4.3Performance evaluations under Weibull Distribution Assumption

Weibull Distributions with scale and shape parameters are assumed as our parametric distribution of the credit scores of the defaulters and the credit scores of the non-defaulters in this performance evaluation. The theoretical value of AUROC for the Weibull score samples is evaluated analytically by the closed form formula (2) in section 3.1 with the location parameters set to zero. The maximum likelihood estimations of sample distribution parameters are obtained by a numerical approximation. Since we have a shape parameter for Weibull Distribution, which may shift the shape of the distribution significantly, we evaluate the performance of the two approaches under two cases: with the same shape parameter for defaulter and non-defaulter sample, and with different shape parameters. The theoretical value of AUROC for the normal score samples is evaluated numerically as well[2]. The rest of the performance evaluation setting is very similar to the Normal Distribution one. There are 6 settings across different AUROC values and defaulter sample sizes as well.

Weibull Setting 1-3:

Normal Distributions			Setting 4	Setting 5	Setting 6
Sample	Shape parameter	Scale	# of Observation	# of Observation	# of Observation
X	2	2	1000	1000	1000
Y	1	1	20	100	500

Theoretical AUROC= 0.757867

Report on estimation error with 50 simulated normal samples & 1000 replication bootstrap

Setting 1

Approach	Non-parametric	Analytic	Difference	Ratio to N
Mean Error	0.005128	0.010230	-0.005102	
Mean ABS Error	0.051701	0.054179	-0.002478	-4.79%
Mean CI Width	0.242836	0.226842	0.015994	6.59%

Setting 2

Approach	Non-parametric	Analytic	Difference	Ratio to N
Mean Error	0.001110	0.000983	0.000127	
Mean ABS Error	0.022661	0.022363	0.000298	1.32%
Mean CI Width	0.112448	0.109910	0.002538	2.26%

Setting 3

Approach	Non-parametric	Analytic	Difference	Ratio to N
Mean Error	0.0027541	0.0030963	-0.000342	
Mean ABS Error	0.0123445	0.0118544	0.000490	3.97%
Mean CI Width	0.0548159	0.0533400	0.001476	2.69%

Weibull Setting 4-6:

Normal Distributions			Setting 4	Setting 5	Setting 6
Sample	Shape parameter	Scale	# of Observation	# of Observation	# of Observation
X	1	3	1000	1000	1000
Y	1	1	20	100	500

Theoretical AUROC= 0.75

Report on estimation error with 50 simulated normal samples & 1000 replication bootstrap

Setting 4

Approach	Non-parametric	Analytic	Difference	Ratio to N
Mean Error	0.000084	0.000314	-0.000231	
Mean ABS Error	0.035960	0.036155	-0.000195	-0.54%
Mean CI Width	0.168248	0.165242	0.003006	1.79%

Setting 5

Approach	Non-parametric	Analytic	Difference	Ratio to N
Mean Error	0.003680	0.003795	-0.000115	
Mean ABS Error	0.018331	0.017988	0.000343	1.87%
Mean CI Width	0.082652	0.081830	0.000822	0.99%

Setting 6

Approach	Non-parametric	Analytic	Difference	Ratio to N
Mean Error	0.003889	0.003961	-0.000072	
Mean ABS Error	0.009632	0.009525	0.005340	1.11%
Mean CI Width	0.048446	0.047586	0.000860	1.77%

In tables 1-6, all mean confidence interval widths show that the estimates of the analytic approach are marginally better then the non-parametric estimates. As for the mean error and the mean absolute error, analytic estimates marginally outperform the non-parametric estimates in tables 2, 3, 5 and 6. Because we use numerical approximation for sample maximum likelihood estimates and the estimation error could be fairly large when we have a small sample, we observe this estimation error is passed through our analytic estimation for the AUROC index, which made the mean absolute errors estimated from the analytic approach are large than the non-parametric approach in setting 1 and 3. This also reduces the gain of the analytic approach over the non-parametric approach in the Weibull Distribution case when comparing with the previous tests.

Summary:

Although the analytic approach gives no better estimates than non-parametric one when we use approximated maximum likelihood estimates for small samples, the performance evaluation shows that the analytic approach works at least as well as the non-parametric approach in above tests and provides better estimate in terms of mean absolute error estimates and confidence interval estimates in most cases.

The above discussion has the following implications. If one can identify some appropriate parametric distributions to the scores of defaulter and non-defaulter, then one could estimate the AUROC and its confidence interval more accurately by the analytic approach. On the other hand, if we can design the rating model so that the score sample is generated by some specific parametric distribution families, then we are able to find a better rating model by using the analytic AUROC as the objective function to maximize in the model selection process.

Another interesting experiment which has not been conducted in previous research is the effect of defaulter's sample size on AUROC. The above experiments clearly show the level of estimation error in both methods with different sample sizes and the resulting error can be large if we only have a small defaulter sample.

In addition, although not very clear but from the results in section 4.1 and 4.2, the analytic approach provides more gain over non-parametric approach when AUROC index is in its high value region than low value region. The reason for this is not clear, we hope to investigate this in future research.

5 Conclusions

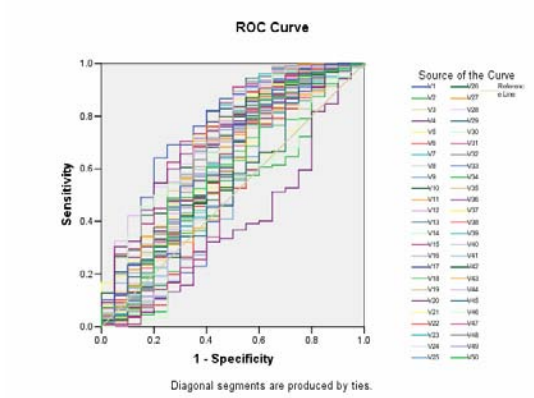
This paper reviewed some of the prevailing credit rating model validation approaches and, in particular, studied the analytic properties of the ROC curve and its summary index AUROC. We use the concept of the population ROC curve to build analytic models of ROC curves. It has been shown through simulation studies that greater accuracy of estimation relative to the non-parametric methods can be achieved. We also show that there are some situations that the accuracy gain of analytic ROC model may decrease, which should be taken into account when applying the analytic models to practical applications.

Moreover, with some particular distributions, where the closed form solution of AUROC is available, analytic AUROC can be directly used as an objective function to maximize during the rating model selection procedure. This means if we can transform the rating scores into those distributions, analytic AUROC could offer a powerful model selection tool.

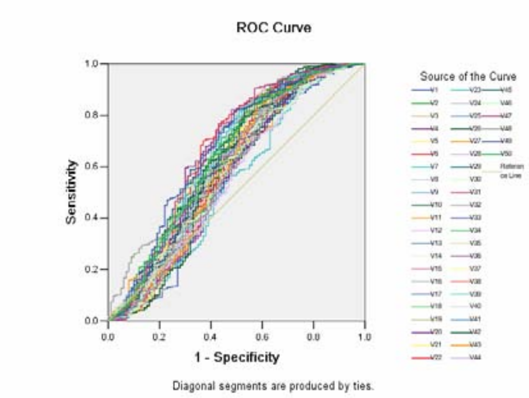
Finally, we also studied the performance of both non-parametric and analytic ROC models under different assumptions for the sample size of defaulters.. The magnitude of error size can be significant when we have a small sample on defaulters, which is a frequently met situation in corporate credit risk study and in periods of benign macro-economic background.

Appendix A1: Non-parametric ROC curve for Normally Distributed Samples

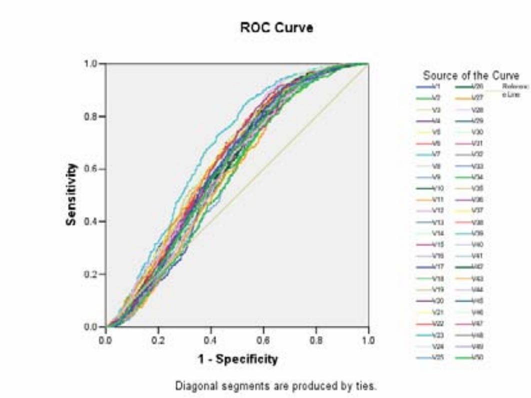
Non-parametric ROC curve under setting 1:



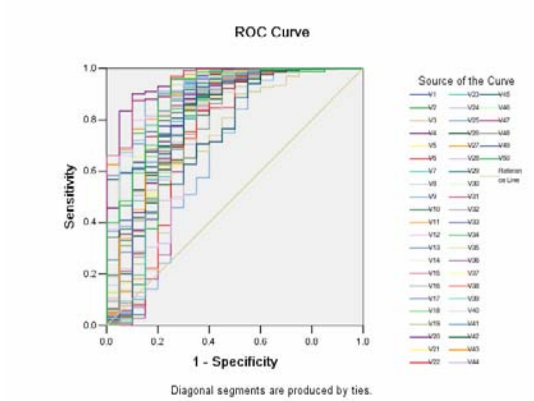
Non-parametric ROC curve under setting 2:



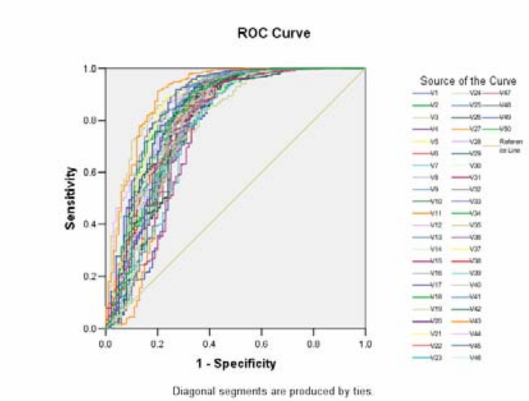
Non-parametric ROC curve under setting 3:



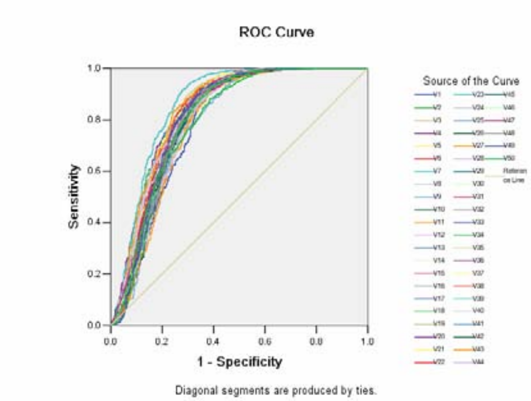
Non-parametric ROC curve under setting 4:



Non-parametric ROC curve under setting 5:



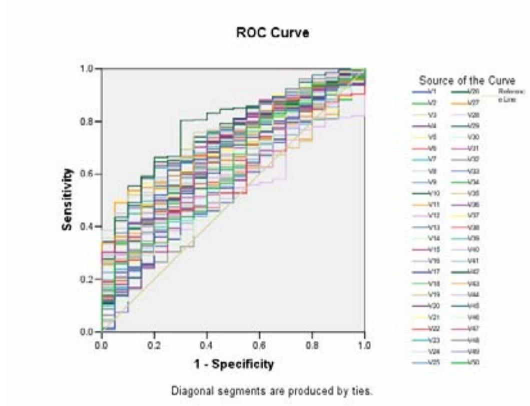
Non-parametric ROC curve under setting 6:



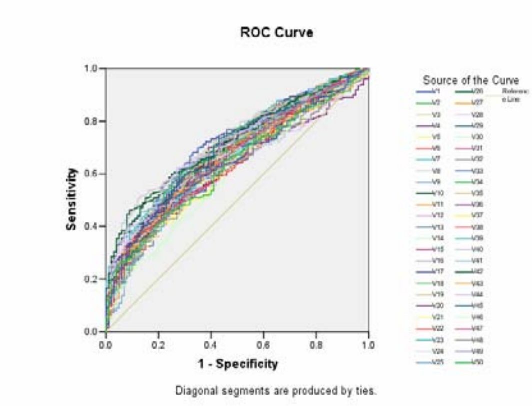
For example, ROC curve V*i* is plotted using the data of simulated sample number *i* generated under a specified setting.

Appendix A2: **Non-parametric ROC curve for Exponentially Distributed Samples**

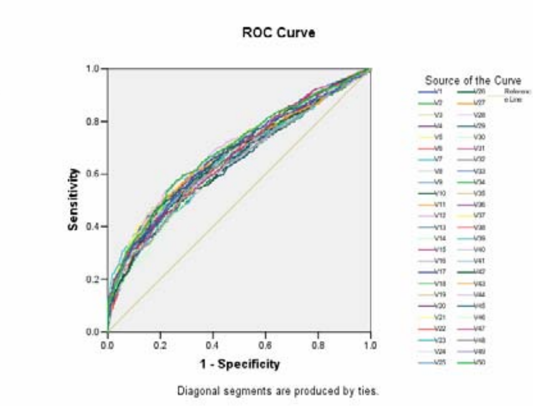
Non-parametric ROC curve under setting 1:



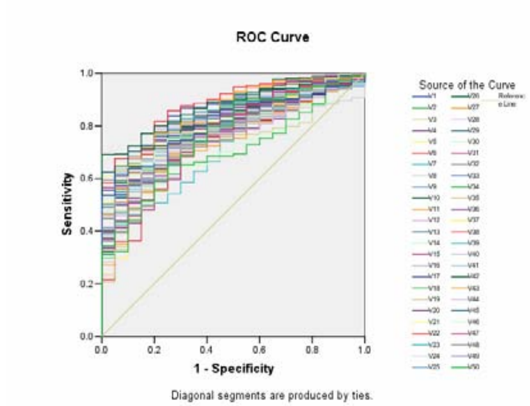
Non-parametric ROC curve under setting 2:



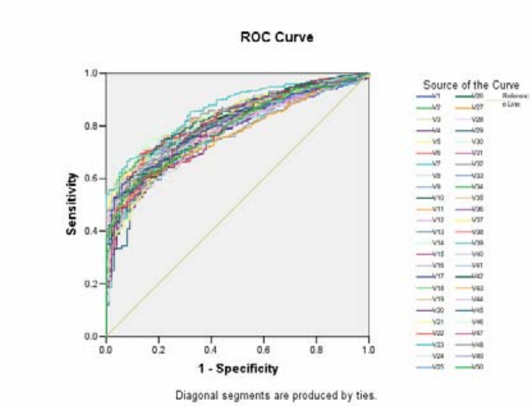
Non-parametric ROC curve under setting 3:



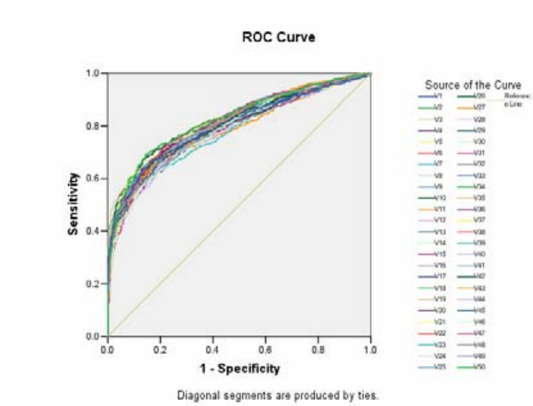
Non-parametric ROC curve under setting 4:



Non-parametric ROC curve under setting 5:



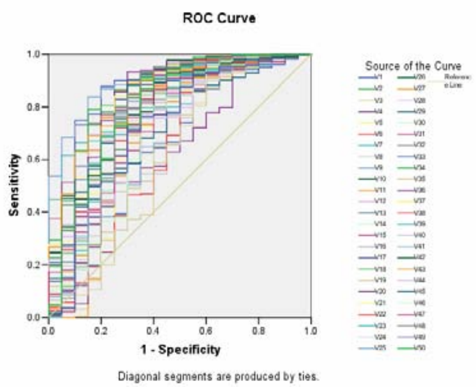
Non-parametric ROC curve under setting 6:



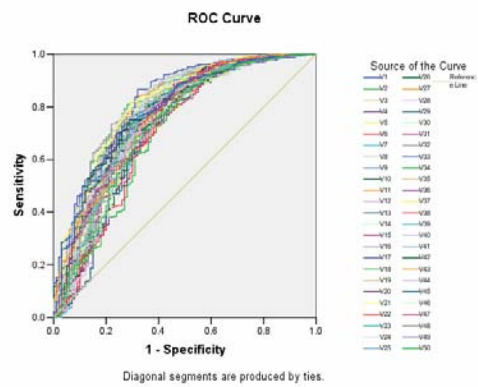
For example, ROC curve V*i* is plotted using the data of simulated sample number *i* generated under a specified setting.

Appendix A3: **Non-parametric ROC curve for Weibull Distributed Samples**

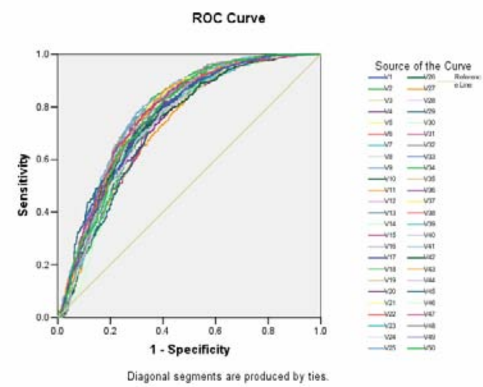
Non-parametric ROC curve under setting 1:



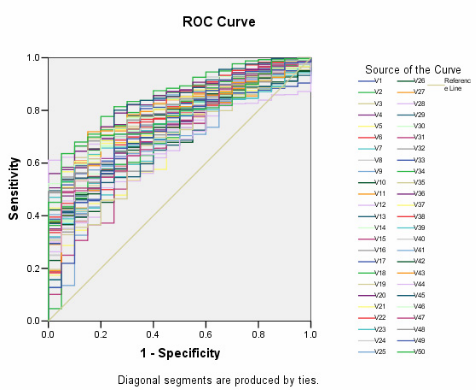
Non-parametric ROC curve under setting 2:



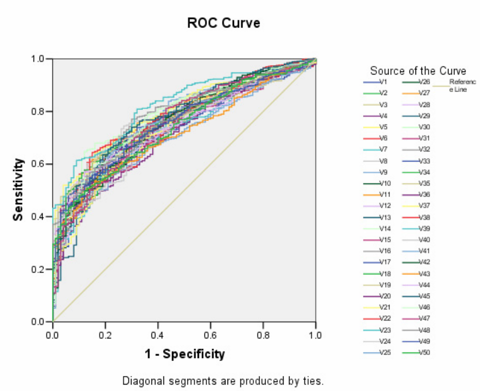
Non-parametric ROC curve under setting 3:



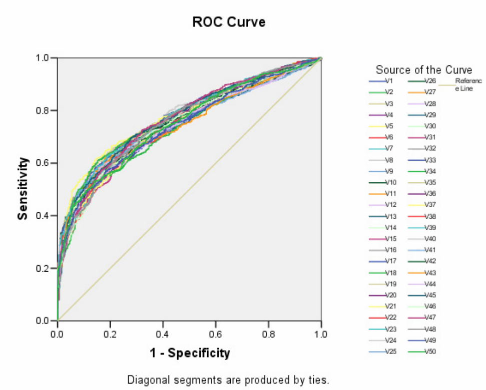
Non-parametric ROC curve under setting 4:



Non-parametric ROC curve under setting 5:



Non-parametric ROC curve under setting 6:



For example, ROC curve V*i* is plotted using the data of simulated sample number *i* generated under a specified setting.

References:

Bamber, Donald (1975), The Area above the Ordinal Dominance Graph and the Area below the Receiver Operating Characteristic Graph. *Journal of Mathematical Psychology* 12, p.387–415.

Basel Committee on Banking Supervision (BCBS) (2004), International Convergence of Capital Measurement and Capital Standards: A Revised Framework. Bank for International Settlements, June.

Basel Committee on Banking Supervision (BCBS) (2005) Working paper No.14

Blochwitz, Stefan, Stefan Hohl and Carsten Wehn (2003), Reconsidering Ratings. Unpublished Working Paper.

Engelmann, Bernd, Evelyn Hayden, and Dirk Tasche, (2003), Testing Rating Accuracy. *Risk*, January 2003, p. 82–86.

Sobehart, Jorge R and Sean C Keenan (2001), Measuring Default Accurately, *Risk Magazine*, March 2001, p. 31-33.

Sobehart, Jorge R and Sean C Keenan (2004), Performance Evaluation for Credit Spread and Default Risk Models. In: Credit Risk: Models and Management. David Shimko (ed.). Second Edition. Risk Books: London, p. 275-305.

Sobehart JR, Sean C Keenan and R Stein (2000), Validation methodologies for default risk models, Credit 1(4), p. 51-56

Swets JA, (1988), Measuring the accuracy of diagnostic systems, Science 240, p. 1285-1293

---

① The theoretical AUROC is approximated by 100000 partitions, while the bootstrap estimation is approximated by 10000 partitions.

② The theoretical AUROC is approximated by 100000 partitions, while the bootstrap estimation is approximated by 10000 partitions.