Assessing the Accuracy of Credit R.O.C. Estimates in the Presence of Macroeconomic Shocks

George A Christodoulakis[1] Stephen E. Satchell[2]

29 May, 2006

Abstract

The Receiver Operating Characteristic (ROC) curve is often used by creditors to assess credit scoring accuracy and as part of their Basel II model validation. The purpose of this paper is to provide a mathematical procedure to assess the accuracy of ROC curve estimates for credit defaults in the presence of macroeconomic shocks. Our approach supplements the non-parametric method recommended by Engelmann et al (2003) based on the Mann-Whitney test which is used as a summary statistic of R.O.C. curves. Our method assumes initially that both sick and healthy loan credit rating scores are generated by normal distributions and shows how R.O.C. estimated depend on the location and scale parameters. We then use these results to construct R.O.C. confidence intervals in closed form and examine the influence of exogenous macroeconomic shocks. We further generalise our method y allowing credit rating scores be generated by skew-normal distributions, thus allowing skewness to affect the moments of the distribution. We show how the presence of skewness would further exacerbate the accuracy of model validation.

[1] Bank of Greece, 21, El. Venizelos Avenue, Athens 10250, gchristodoulakis@bankofgreece.gr and Manchester Business School, University of Manchester, Booth Street West, Manchester M15 6PB, UK. The views expressed in this paper are those of the authors and should in no part be attributed to the Bank of Greece

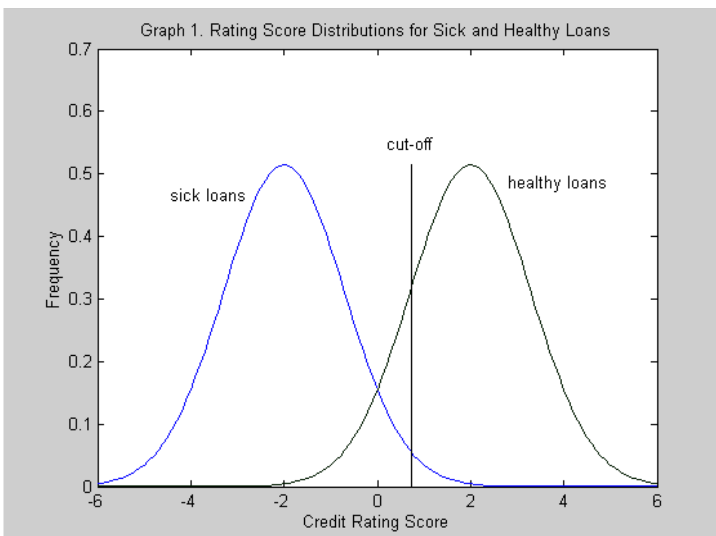
[2] University of Cambridge, Trinity College and University of London, Birkbeck College and Visiting Advisor at the Bank of Greece, ses11@econ.cam.ac.uk

1. Introduction

The validation of credit risk models lies at the heart of credit risk management processes developed in financial institutions in the context of Basel II recommendations. The quality of models is often judged on the basis of statistical metrics of discriminatory power as well as default forecasting ability. The Receiver Operating Characteristic (ROC) and the Cumulative Accuracy Profile (CAP) curves as well as their summary statistics of Accuracy Ratio (AR) and Area Under R.O.C. (AUROC) are often used by creditors to assess credit scoring accuracy and as part of their Basel II model validation. Related tests but with relative limited applications are the Kolmogorov-Smirnov test, the Mahalanobis distance as well as the Gini coefficient. Recent review papers have been published by Tasche (2005) and Christodoulakis (2006). The purpose of this paper is to provide a mathematical procedure to assess the accuracy of ROC curve estimates for credit defaults in the presence of macroeconomic shocks. Our approach supplements the non-parametric method recommended by Engelmann et al (2003) based on the Mann and Whitney (1947) test. Our method assumes initially that both sick and healthy loan credit rating scores are generated by normal distributions and shows how R.O.C. estimated depend on the location and scale parameters. We then use these results to construct R.O.C. confidence intervals in closed form and examine the influence of exogenous macroeconomic shocks. We further generalise our method by allowing credit rating scores be generated by skew-normal distributions, thus allowing skewness to affect the moments of the distribution. We show how the presence of skewness would further exacerbate the accuracy of model validation.

2. A Model for ROC Confidence Intervals

Consider two absolutely continuous random variables, y and x, that refer to sick and healthy credit scores respectively. Their distribution functions are defined as $F_y(c) = \Pr(y \le c)$ and $F_x(c) = \Pr(x \le c)$ respectively and $F_x^{-1}(.)$ denotes the inverse distribution function which is uniquely defined. A perfect rating model should completely separate the two distributions whilst for an imperfect (and real) model perfect discrimination is not possible and the distributions should exhibit some overlap. The latter situation is presented in Graph 1 using normal density functions.



A natural way for a decision maker to discriminate the debtors belonging to the two classes is to introduce a cut-off point as depicted in Graph 1, which would classify all the debtors below that point as potential defaulters and those above as potential survivors. This practice introduces four possible decision results as clearly described by Sobehart and Keenan (2001): (1) debtors classified bellow cut-off which eventually defaulted (Correct Alarms), (2) debtors classified bellow cut-off which eventually survived (False Alarms), (3) debtors classified above cut-off which eventually survived (Correct Survivors) and (4) debtors classified above cut-off which eventually defaulted (Missed Alarms).

The Receiver Operating Characteristic (R.O.C.) is constructed by calculating for every possible cut-off point in the range of rating scores, the ratio of Correct Alarms to total number of defaults (Correct Alarm Rate (CAR)) and False Alarms to total number of non-defaults (False Alarm Rate (FAR)). Then, R.O.C. is defined as the plot of pairs of CAR versus FAR. Clearly, both quantities take values between zero and one and in Graph 1 CAR can be represented by the integral of the sick loan density up to the cut-off point whilst FAR can be represented by the integral of the healthy loan density up to the cut-off point. This probabilistic interpretation leads us to state the following proposition.

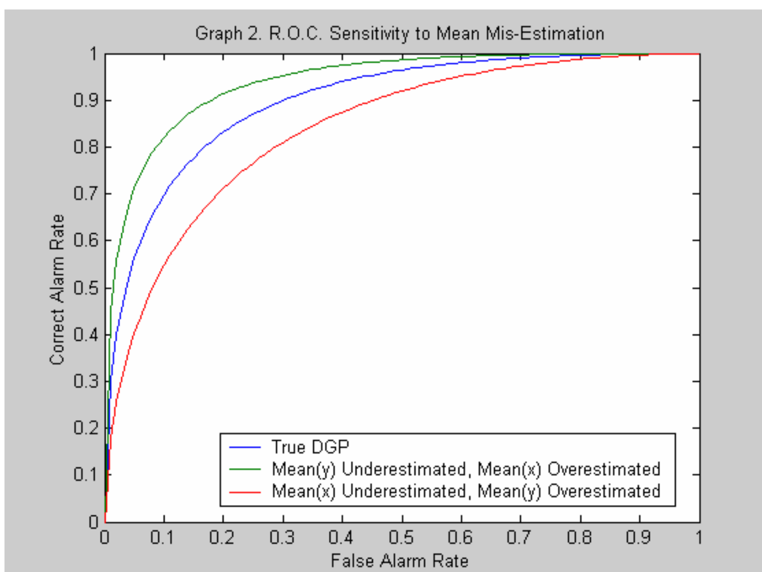
Proposition 1. If the credit rating scores for defaulters *y* and non-defaulters *x* are represented by mutually independent normally distributed random variables $y \sim N\left(\mu_y, \sigma_y^2\right)$ and $x \sim N\left(\mu_x, \sigma_x^2\right)$ respectively, then the Receiver Operating Characteristic satisfies the following relationship.

$$CAR = \Phi\left(\Phi^{-1}(FAR)\right) = \Phi\left(\frac{\left(\mu_x - \mu_y\right) + \Phi^{-1}(FAR)\sigma_x}{\sigma_y}\right)$$

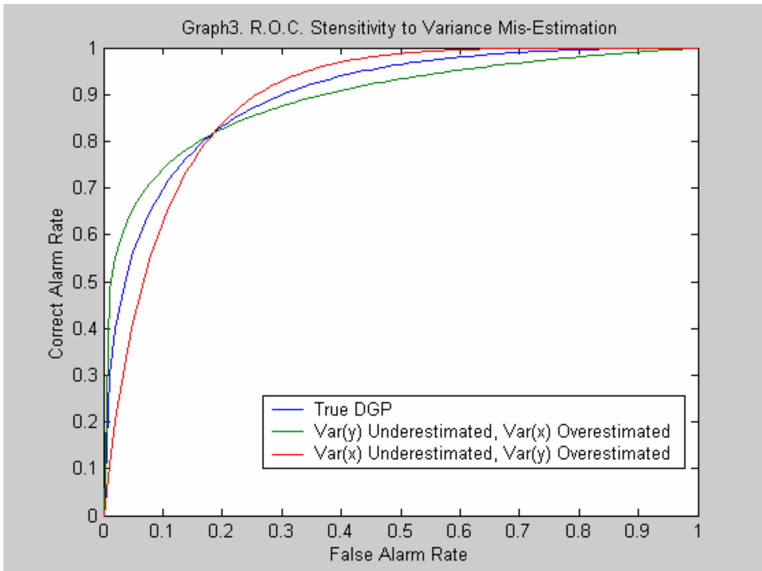
where Φ( ) denotes the cumulative standard normal density.

Proof: See Appendix.

Given two samples $\left(x_1, \ldots, x_{N_1}\right)$ and $\left(y_1, \ldots, y_{N_2}\right)$ we can estimate the unknown parameters μ_x, μ_y, σ_x, σ_y, as the usual sample moments. There may be an alternative minimum variance unbiased estimator for equation (1) given the completeness and efficiency of estimates $\bar{x}, \bar{y}, s_x^2, s_y^2$ and the results of the Rao-Blackwell theorem; we shall return to this point at a later date. The relevance of estimated parameters in our analysis can be illustrated by considering a numerical example. Consider a "true" data generating process in which the means of sick and healthy loan credit rating scores are -7 and 2 respectively and their standard deviation is 5. Let us assume that sick (healthy) loan mean has been underestimated (overestimated) taking value -8 (3). This would result in a false belief that the rating model exhibits superior performance over the entire range of false alarm rates. The reverse results would become obvious in the case that sick (healthy) loan mean had been overestimated (underestimated) taking value -6 (1). We plot all three cases in Graph 2.



Then, let us assume that sick (healthy) loan standard deviation has been underestimated (overestimated) taking value 4 (7). This would result in a false belief that the rating model exhibits superior (inferior) performance for low (high) false alarm rates. The reverse results would become obvious in the case that sick (healthy) loan standard deviation had been overestimated (underestimated) taking value 7 (4). We plot all three cases in Graph 3.



Turning now our attention to the construction of R.O.C. confidence intervals, for the moment, we denote our estimated y by $CAR(\hat{\theta}, FAR)$ versus the true $CAR(\theta, FAR)$, where $\theta = (\mu_x, \mu_y, \sigma_x, \sigma_y)$. Taking Taylor series expansion we can write

$$CAR(\hat{\theta}, FAR) - CAR(\theta, FAR) = \left.\frac{\partial CAR}{\partial \theta}\right|_{\hat{\theta}}^{'} (\hat{\theta} - \theta) \qquad (2)$$

and compute the asymptotic distribution of $CAR(\hat{\theta}, x)$ to arrive at a confidence interval for it. Noting that

$$\sqrt{M}(\hat{\theta} - \theta) \xrightarrow{d} N(\mathbf{0}, \Omega_{\theta\theta})$$

where $M = \min(N_1, N_2)$, we see that

$$\sqrt{M}\left(CAR(\hat{\theta}, x) - CAR(\theta, x)\right) \xrightarrow{d} N\left(0, \left.\frac{\partial CAR}{\partial \theta}\right|_{\theta}^{'} \Omega_{\theta\theta} \left.\frac{\partial CAR}{\partial \theta}\right|_{\theta}\right) \tag{3}$$

In what follows, we describe $\frac{\partial CAR}{\partial \theta}$, $\Omega_{\theta\theta}$ under our assumption of joint normality. Let

$$\begin{aligned} CAR(\hat{\theta}, FAR) &= \Phi\left(\frac{(\hat{\mu}_x - \hat{\mu}_y) + \Phi^{-1}(FAR)\hat{\sigma}_x}{\hat{\sigma}_y}\right) \\ &= \Phi(\hat{a}) \end{aligned} \tag{4}$$

Now

$$\frac{\partial \Phi(a)}{\partial a} = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{a^2}{2}\right) = \varphi(a)$$

and

$$\left.\frac{\partial a}{\partial \theta}\right|_{\hat{\theta}=\theta} = \left(\frac{1}{\sigma_y}, -\frac{1}{\sigma_y}, \frac{\Phi^{-1}(x)}{\sigma_y}, -\frac{(\mu_x - \mu_y) + \Phi^{-1}(x)\sigma_x}{\sigma_y^2}\right)$$

Also, $\Omega_{\theta\theta}$ is diagonal with elements $\left(\sigma_x^2, \sigma_y^2, \frac{\sigma_x^2}{2}, \frac{\sigma_y^2}{2}\right)$ where the last two have been calculated by the delta method. We could, however, use the exact formula for them if needed. Completing the matrix calculations in (3), we see that any variance of $\sqrt{M}\left(CAR(\hat{\theta}, FAR) - CAR(\theta, FAR)\right)$ is given by

$$
\begin{aligned}
d^2 &= \frac{\varphi(a)^2}{\sigma_y^2}\begin{pmatrix} 1 \\ -1 \\ h \\ -a \end{pmatrix}' \begin{pmatrix} \sigma_x^2 & & & \\ & \sigma_y^2 & & \\ & & \frac{\sigma_x^2}{2} & \\ & & & \frac{\sigma_y^2}{2} \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ h \\ -a \end{pmatrix} \\
&= \frac{\phi^2(a)}{2\sigma_y^2}\left(2\sigma_x^2 + 2\sigma_y^2 + h^2\sigma_x^2 + a^2\sigma_y^2\right) \qquad (5) \\
&= \frac{\exp\left(-a^2\right)}{\pi\sigma_y^2}\left(2\sigma_x^2 + 2\sigma_y^2 + h^2\sigma_x^2 + a^2\sigma_y^2\right)
\end{aligned}
$$

where

$$
\begin{aligned}
h &= \Phi^{-1}(x) \\
a &= \frac{\left(\mu_x - \mu_y\right) + h\sigma_x}{\sigma_y}
\end{aligned}
$$

It is now possible to state a simple proposition for the confidence interval of CAR.

Proposition 2. If the credit rating scores for defaulters *y* and non-defaulters *x* are represented by mutually independent normally distributed random variables $y \sim N\left(\mu_y, \sigma_y^2\right)$ and $x \sim N\left(\mu_x, \sigma_x^2\right)$ respectively and Proposition 1 holds, then the confidence interval for CAR is given by

$$
CAR\left(\hat{\theta}, FAR\right) - \Phi^{-1}\left(\frac{1+alpha}{2}\right)d < CAR\left(\theta, FAR\right) < CAR\left(\hat{\theta}, FAR\right) + \Phi^{-1}\left(\frac{1+alpha}{2}\right)d \qquad (6)
$$

where $\Phi^{-1}\left(\frac{1+alpha}{2}\right)$ is the upper *alpha* per cent point of the standard normal.

Proof: The proof is outlined in the analysis preceding the proposition □

The width of the confidence interval is given by $2\Phi^{-1}\left(\frac{1+alpha}{2}\right)d$ and changes in *FAR* change h and a through d via equation (5). Recall that we have assumed that $M=\min(N_1,N_2)\to\infty$. A more precise formulation which reflects the much smaller number of sick loans versus healthy ones is that $N_1=\lambda N_2$ and keep $0<\lambda<1$ fixed whilst $N_2\to\infty$. The impact of this is to replace σ_y^2 from equation (3) onwards by $\frac{\sigma_y^2}{\lambda}$. In this section we have relied on the assumption of normality as a data generating mechanism for the credit rating scores of defaulters and non-defaulters. Although empirically relevant, this is a rather simplifying assumption which serves the goals of analytical and intuitive tractability and in Section 4 we shall extend our analysis to the case of non-normality.

3. Incorporating Macroeconomic Shocks

Suppose that our random variables are conditioned on a random variable $z\sim(\mu_z,\sigma_z^2)$, possibly macroeconomic, so that

$$\begin{aligned}x|z &\sim N\left(\mu_x^{'}+\beta_x(z-\mu_z),\sigma_x^{'2}\left(1-\rho_{xz}^2\right)\right)\\ y|z &\sim N\left(\mu_y^{'}+\beta_y(z-\mu_z),\sigma_y^{'2}\left(1-\rho_{yz}^2\right)\right)\end{aligned}$$

Interpreting our earlier calculation and now regarding our original parameters as

$$\begin{aligned}\mu_x &= \mu_x^{'}+\beta_x(z-\mu_z)\\ \mu_y &= \mu_y^{'}+\beta_y(z-\mu_z)\\ \sigma_x^2 &= \sigma_x^{'2}\left(1-\rho_{xz}^2\right)\\ \sigma_y^2 &= \sigma_y^{'2}\left(1-\rho_{yz}^2\right)\end{aligned}$$

We have assumed that $\operatorname{cov}(x|z, y|z) = 0$, however

$$\begin{aligned}\operatorname{cov}(x, y) &= E_z(\operatorname{cov}(x|z, y|z)) + \operatorname{cov}(E(x|z), E(y|z)) \\ &= \beta_x \beta_y \sigma_z^2\end{aligned}$$

Note that

$$\mu_x - \mu_y = \mu_x^{'} - \mu_y^{'} + (\beta_x - \beta_y)(z - \mu_z)$$

This point to the obvious intuition that the ROC curve will be invariant to changes in *z* if the impact of healthy loans β_x is the same as the impact of sick loans β_y. If *z* represents the level of interest rate, we might expect that $\beta_y < \beta_x < 0$ so that an increase in interest rates should increase $\mu_x - \mu_y$ and sharpen the test's discriminative power at least at some levels of *x*. This result is relevant to the issue, often discussed by practitioners, that the current macroeconomic environment is too benign to allow discrimination between sick and healthy mortgages, especially in a world with near-zero defaults. In what follows we trace through the effect of an increase in $\mu_x - \mu_y$ and the width of the asymptotic confidence interval described in equations (5) and (6). It is clear that changes in $\mu_x - \mu_y$ only influence *a*. Furthermore, the latter only appears as a^2 so it is absolute *a* that matters. Note also that if $x < 0.5$, $h < 0$ so that an increase in $\mu_x - \mu_y$ will reduce a^2 over a certain range and thus reduce d^2, making our model discrimination more precise. For a change Δ in $\mu_x - \mu_y$, a^2 will be reduced if $\mu_x - \mu_y < -h - \frac{\Delta}{2}$ where $h < 0$, $\Delta > 0$. Now $\Delta = (\beta_x - \beta_y)\Delta z$, for $\Delta z > 0$ meaning the change in the conditioning variable.

4. Generalising to Non-Normal Families of Densities

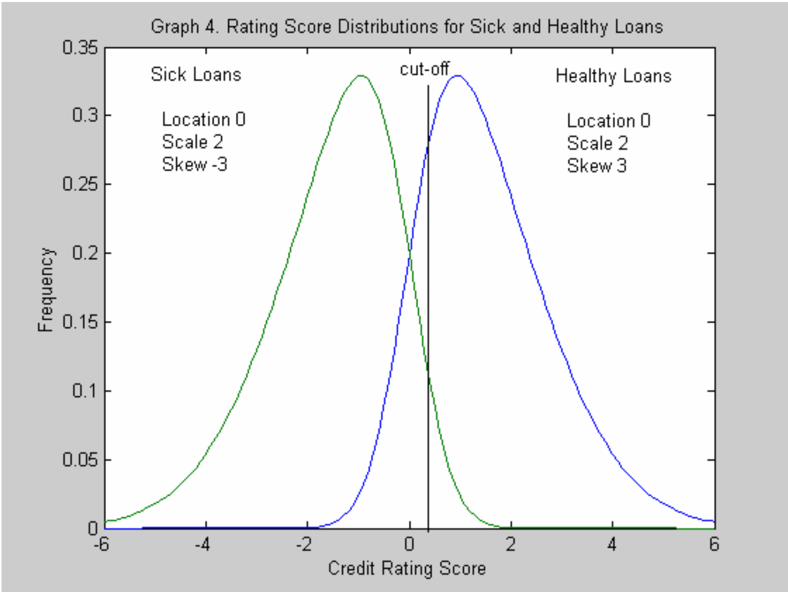
In Section 2 we have assumed normality as a data generating mechanism for the credit rating scores of defaulters and non-defaulters. We shall study the effects of non-normality by assuming that our data are generated by independent Skew-Normal distributions, originally introduced by O'Hagan and Leonard (1978) as priors in Bayesian estimation and developed by Azzalini (1985, 1986) and further generalised by Azzalini and Dalla Valle (1996) and Arnold and Lin (2004) among others. Let $y = \mu_y + \sigma_y v_y$ where $v_y \sim SN(\lambda_y)$ and λ_y is a real constant then the density function of the Skew Normal distribution for v_y is given by

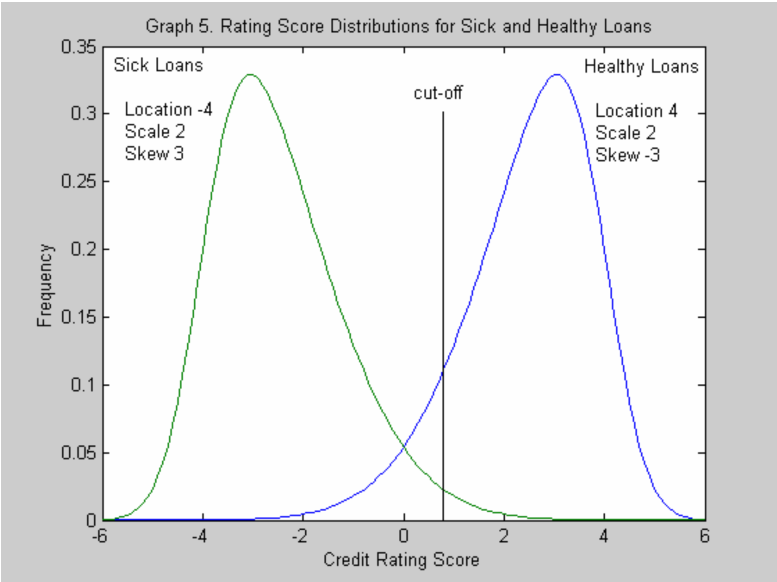
$$pdf(v_y) = 2\varphi(v_y)\Phi(\lambda_y v_y)$$

where

$$\varphi(v_y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{v_y^2}{2}\right)$$

is the standard normal density and $\Phi(\ )$ is the cumulative standard normal. The Skew Normal accommodates a variety of skewness patterns as λ varies, whilst it converges to the Normal as $\lambda \to 0$. Similarly, for non-default data we assume $x = \mu_x + \sigma_x v_x$ where $v_x \sim SN(\lambda_x)$. However, in this case μ, σ, and λ refer to location, scale and skewness parameters respectively and do not correspond to moments. Instead, we can show that $E(x) = \mu + \sigma\lambda\sqrt{\frac{2\pi}{1+\lambda^2}}$, $Var(x) = \sigma^2\left(1 - \frac{2\lambda^2}{\pi(1+\lambda^2)}\right)$ whilst both skewness and kurtosis also depend on λ. Thus, in the presence of Skew-Normal data generating processes, the decision problem of discriminating between default and non-default scoring distributions, as depicted in Graph 1, would have a large number of analogues depending on the relative values of skewness parameters λ_x and λ_y on top of location and scale parameters. As an example, for $\lambda_y = 3$ and $\lambda_x = -3$ the likelihood of making discrimination errors is shown to decrease in Graph 4, but when $\lambda_y = -3$ and $\lambda_x = 3$ we observe clearly that the distributions develop extensive overlap which in turn enhances the likelihood of making both types of discrimination errors, see Graph 5.





Our results under Skew-Normal are now summarised in Proposition 3.

Proposition 3. If the credit rating scores for defaulters *y* and non-defaulters *x* are represented by mutually independent Skew-Normally distributed random variables $y \sim SN\left(\mu_y, \sigma_y^2, \lambda_y\right)$ and $x \sim SN\left(\mu_x, \sigma_x^2, \lambda_x\right)$ respectively, then the Receiver Operating Characteristic satisfies the following relationship.

$$
\begin{aligned}
CAR &= CSN\left(\frac{\left(\mu_x - \mu_y\right) + CSN^{-1}\left(FAR; \lambda_x\right)\sigma_x}{\sigma_y}; \lambda_y\right) \\
&= \Phi\left(\frac{\left(\mu_x - \mu_y\right) + CSN^{-1}\left(FAR; \lambda_x\right)\sigma_x}{\sigma_y}\right) - 2T\left(\frac{\left(\mu_x - \mu_y\right) + CSN^{-1}\left(FAR; \lambda_x\right)\sigma_x}{\sigma_y}; \lambda_y\right)
\end{aligned}
$$

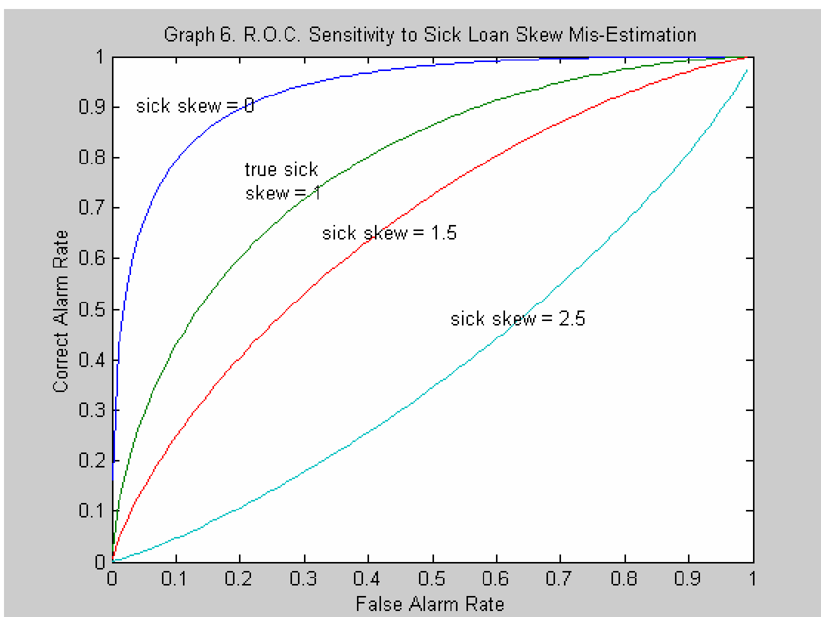
where CSN( ) and Φ( ) denote the cumulative Skew-Normal and Standard Normal densities respectively and T( ) denotes the Owen (1956) function.

Proof: We follow the same steps as in Proposition 1 under Skew-Normality and the result (f) of Azzalini (2005) that CSN(w;k) = Φ(w) – 2T(w;k). □

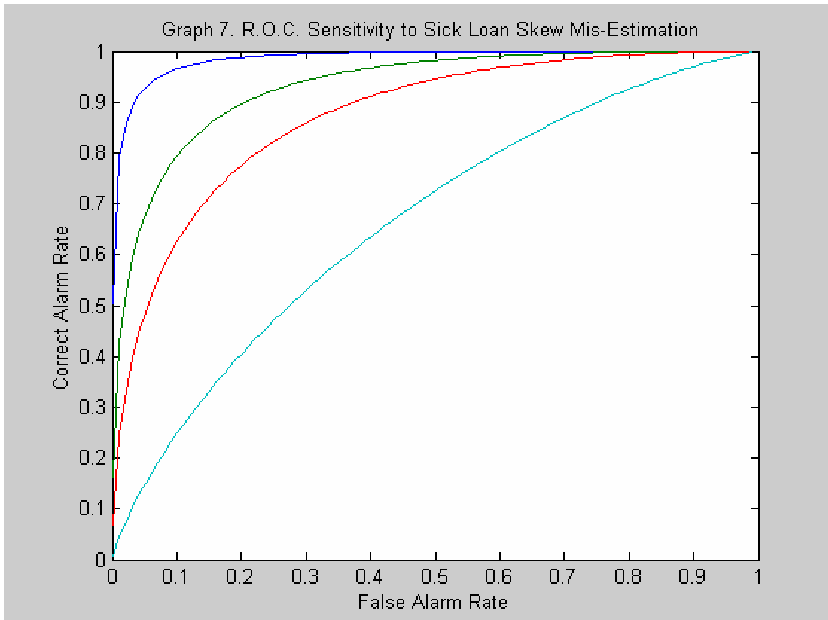
Note that given values of g and k, the Owen (1956) function T(w.k) calculates the quantity

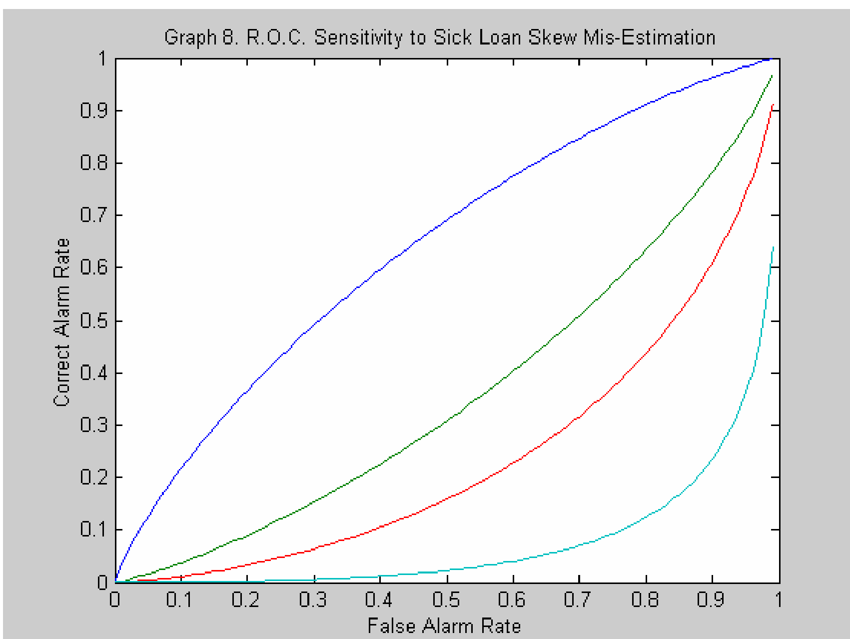
$$
T(w, k) = \frac{1}{2\pi}\int_0^k \frac{\exp\left(-\frac{w^2}{2}\left(1 + x^2\right)\right)}{1 + x^2} dx
$$

The ROC curve described in Proposition 3 has a more general form as compared with that of Proposition 1 in that it is affected not only by location and scale parameters but also by skewness. This allows for further flexibility and accuracy in generating ROC curves as we can show that the four moments of the skew-normal distribution are all affected by the presence of skewness. Consider "true" data generating process in which the means of sick and healthy loan credit rating scores are -7 and 2 respectively and their standard deviation is 5. Also the true sick and healthy loan skewness parameters are 1 and 0.3 respectively. Let us assume that sick loan skew parameter has been mis-estimated taking possible values 0, 1.5 and 2.5 respectively. We plot these alternative R.O.C. curves in Graph 6 and we observe clearly that sick skewness under-estimation (over-estimation) results in a false belief of rating model superior (inferior) performance over the entire range of false alarm rates. Ultimate under-(over) estimation of skewness parameter would lead the analyst to the false conclusion that the model approaches perfectly efficient (inefficient) performance.



These comparative static effects would be effectively altered in the case both sick and healthy loan parameters mis-estimation or under different true data generating processes. For example, using all the parameter values as described above but for healthy loan skewness 1.3 or -1.3, our R.O.C. results would be depicted as in Graph 7 and 8 respectively.





Graphs 7 and 8 show that our false impression on the performance of a model is subject to skew parameter trade-offs between sick and healthy loan distributions, whilst the picture would further complicate if location and scale parameters change as well. The relationship between Proposition 1 and Proposition 3 can be described in the following corollary.

Corollary. The Receiver Operating Characteristic of Proposition 1 is a special case of Proposition 3.

Proof. Letting $\lambda_x \to 0$ and $\lambda_y \to 0$, the skew-normal densities reduce to standard normal and thus the R.O.C. curve of Proposition 3 reduces to Proposition 1.□

5. Concluding Remarks

The Receiver Operating Characteristic (ROC) curve is often used by creditors to assess credit scoring accuracy and as part of their Basel II model validation. The purpose of this paper is to provide a mathematical procedure to assess the accuracy of ROC curve estimates for credit defaults in the presence of macroeconomic shocks. Our approach supplements the non-parametric method recommended by Engelmann et al (2003) based on the Mann and Whitney (1947) test which is used as a summary statistic of R.O.C. curves. Our method assumes initially that both sick and healthy loan credit rating scores are generated by normal distributions and shows how R.O.C. estimated depend on the location and scale parameters. We then use these results to construct R.O.C. confidence intervals in closed form and examine the influence of exogenous macroeconomic shocks. We further generalise our method y allowing credit rating scores be generated by skew-normal distributions, thus allowing skewness to affect the moments of the distribution. We show how the presence of skewness would further exacerbate the accuracy of model validation. Future work would consider the implications of our approach to the construction of confidence intervals under non-normality as well as the power of bootstrap methods.

Appendix

Proof of Proposition 1

Let v_j denote standard normal innovations so that we can write

$$y = \mu_y + \sigma_y v_y$$
$$x = \mu_x + \sigma_x v_x$$

Then by definition we have

$$CAR = \int^{c} f(y)dy = \Phi\left(\frac{c-\mu_y}{\sigma_y}\right)$$
and
$$FAR = \int^{c} f(x)dx = \Phi\left(\frac{c-\mu_x}{\sigma_x}\right)$$

Solving both equations with respect to *c* and equating we obtain

$$\Phi^{-1}(CAR)\sigma_y + \mu_y = \Phi^{-1}(FAR)\sigma_x + \mu_x$$

leading to

$$CAR = \Phi\left(\frac{(\mu_x - \mu_y) + \Phi^{-1}(FAR)\sigma_x}{\sigma_y}\right)$$

□

References

Arnold, B. C. and G. D. Lin (2004), Characterizations of the Skew-Normal and Generalized Chi Distributions, *Sankhyã: The Indian Journal of Statistics*, Vol. 66, (4), pp. 593-606

Azzalini, A. (1985), A Class of Distributions which Includes the Normal Ones, *Scandinavian Journal of Statistics*, Vol. 12, pp. 171-178

Azzalini, A. (1986), Further Results on a Class of Distributions which Includes the Normal Ones, *Statistica*, Vol. 46, pp. 199-208

Azzalini, A. and A. Dalla Valle (1996), The Multivariate Skew-Normal Distributions, Biometrika, Vol. 83, pp. 715-726

Christodoulakis, G. A. (2006), The Validity of Credit Risk Model Validation Methods, forthcoming in Christodoulakis, G. A. and S. E. Satchell (eds.), *The Analytics of Risk Model Validation*, Elsevier-Butterworth-Heinemann

Engelmann B., E. Hayden and D. Tasche, (2003), Testing Rating Accuracy, *Risk*, Vol. 16, January, pp. 82-86

Mann H. and D. Whitney (1947), On a Test of Whether One or Two Random Variables is Stochastically Larger than the Other, *Annals of Mathematical Statistics*, Vol. 18. pp. 50-60

O'Hagan, A. and T. Leonard (1978), Bayes Estimation Subject to Uncertainty about Parameter Constraints, *Biometrika*, Vol. 63, pp. 201-203

Owen, D. B. (1956), Tables for Computing Bivariate Normal Probabilities, *Annals of Mathematical Statistics*, Vol. 27. 1075-1090

Sobehart, J. and S. Keenan, (2001), Measuring Default Accurately, *Risk*, Vol. 11 (41), pp. S31-S33

Tasche D. (2006), Rating and Probability of Default Validation, in *Studies on the Validation of Internal Rating Systems*, BIS Working Paper No 14., pp. 28-59