ISSN: 2281-1346

Department of Economics and Management

DEM Working Paper Series

Credit risk predictions with Bayesian model averaging

Silvia Figini
(Università di Pavia)

Paolo Giudici
(Università di Pavia)

 # 34 (02-13)

Via San Felice, 5
I-27100 Pavia
http://epmq.unipv.eu/site/home.html

February 2013

Credit risk predictions with Bayesian model averaging

S. Figini and P. Giudici
Department of Economics and Management
University of Pavia

Abstract

Model uncertainty remains a challenge to researchers in different applications. When many competing models are available for estimation, and without enough guidance from theory, model averaging represents an alternative to model selection. Despite model averaging approaches have been present in statistics for many years, only recently they are starting to receive attention in applications.

The Bayesian Model Averaging (BMA) approach sometimes can be difficult in terms of applicability, mainly because of the following reasons: firstly two types of priors need to be elicited and secondly the number of models under consideration in the model space is often huge, so that the computational aspects can be prohibitive.

In this paper we show how Bayesian model averaging can be usefully employed to obtain a well calibrated model, in terms of predictive accuracy for credit risk problems. In this paper we shall investigate how BMA performs in comparison with classical and Bayesian (single) selected models using two real credit risk databases.

1 Motivation

In this paper we concentrate our attention on a general class of parametric classification models aimed at explaining a binary target response variable, a credit default event as a function of a set of explanatory variables.
In the above context, statistical models are usually chosen according to a

model selection procedure that aims at selecting the most performing structure. The chosen model is, once selected, taken as the basis for further actions, such as parameter estimation, default prediction and predictive classification.
However, relying upon a single model may not be the best strategy, as the uncertainty over the model space is not adequately taken into account.
Few papers have investigated the comparison between single selected models and model averaging in credit risk modelling. Among them, we recall the paper of Hayden et al. (2009), which presents a comparison between stepwise selection in logistic regression and Bayasian Model Averaging for credit risk models. Another reference is Tsai et al. (2010) that show a statistical criterion and a financial market measure to compare the forecasting accuracy of different model selection approaches: Bayesian information criterion, model averaging and model mixing.
The main objective of this paper is to extend the previous contributions, comparing the relative performance of single credit risk models with that of models constructed by means of a model averaging approach. Ley and Steel (2007) and Eicher et al. (2011) show that prior assumptions can be extremely critical for the outcome of a model averaging analysis. Here we focus the hierarchical prior analysis of Ley and Steel (2007).

In order to show how our proposal work we compute different kind of measures aimed at measuring the predictive ability, the discriminant power and the selectivity of each model proposed (see e.g. Hand et al. 2010, Figini and Giudici 2011).

The paper is structured as follows: Section 2 describes the background of our proposal; Section 3 reports our proposed model; Section 4 reports and compares the empirical evidences achieved on the simulated and on the real data sets. Finally, conclusions and further ideas of research are reported in Section 5.

2 Background

The common practice in empirical research is selecting a single model after what amounts to be a search in the space of all possible models. Then, researchers typically base their conclusions acting as if the model chosen were

the true model. However, this procedure tends to understate the variability contained in the analysed data, that may support a plurality of competing models. A way to overcome this problem is to use a model averaging approach (see e.g. Hoeting et al. 1999).
The literature proposes different approach to treat model averaging, both in the frequentist and in the Bayesian framework. Here we consider the latter.

Let us suppose that a researcher has q possible models in mind, $h = 1, \dots, q$. This implies that there are q different estimates of the parameter of interest depending on the model considered, say $(\hat{\beta}_1, \dots, \hat{\beta}_q)$. The key idea of model averaging is to consider and estimate all the q candidate models and then report a weighted average as the estimate of the effect of interest. Such model averaging estimate is:

$$\hat{\beta_{MA}} = \sum_{h=1}^{q} \omega_h \hat{\beta}_h, \tag{2.1}$$

where ω_h is the weight associated to model h. Here we focus on the Bayesian approach to model averaging. Given q variables we obtain 2^q possible models, $M_1, \dots, M_{2^q}$. Given an observed data vector y, the posterior distribution for the parameters of model M_j, β^j, can be obtained from:

$$g(\beta^j|y, M_j) = \frac{f(y|\beta^j, M_j)g(\beta^j|M_j)}{f(y|M_j)}. \tag{2.2}$$

Formula (2.2) shows that, for each model M_j, we need to calculate a likelihood $f(y|\beta^j, M_j)$ and a prior distribution $g(\beta^j|M_j)$, from which the posterior distribution $g(\beta^j|y, M_j)$, can be obtained and, then, summarised by the posterior mean $\beta^j = E(\beta^j|y, M_j)$, the optimal estimate under a quadratic loss function (see e.g. Bernardo and Smith 1994).
We have thus obtained the first component of the model averaged estimate in equation 2.1. The second component is the model weight, that can be obtained as the posterior distribution of a model:

$$p(M_j|y) = \frac{f(y|M_j)p(M_j)}{f(y)}, \tag{2.3}$$

where $f(y|M_j)$ is the marginal likelihood of a model j, and $p(M_j)$ is the prior belief on the same model.

From the above premises, note that Bayesian Model Averaging (BMA hence forth) involves two different priors beliefs, one on the parameter space $g(\beta^j|y, M_j)$ and another one on the model space $p(M_j)$.

In this paper we focus on the choice of the prior on the model space, assuming a non informative prior on the parameter space, following the approach of Hoeting et al. (1999). This because our main aim is to compare fairly the predictive capacity of the BMA approach not only with respect to a single chosen Bayesian model, but also with respect to a classical, non Bayesian approach, that can be assimilated to a Bayesian approach with a non informative prior on the parameter space (see e.g. Bernardo and Smith, 1994).

A key aspect of our work is the comparison with (single) selected models with BMA by means of key performance indicators able to measure the predictive capability. In order to detect the *predictive capability* of a model, we employ the confusion matrix and related measures of interest (see e.g. Kohavi et al., 1997). The models will be compared using different cut-offs. The literature proposes optimisation techniques able to derive the best cut-off resorting for example to the minimisation of the difference between sensitivity and specificity (P-fair in Schrder and Richter 1999) or to the maximisation of the correct classification rate (P-opt, calculated from the ROC as described in Zweig and Campbell, 1993) taking into account different costs of false positive or false negative predictions). In our example we fix, for the sake of comparison, a cut-off =0.5 and a cut-off equal at the observed frequency of the event of interest, the credit default in our case.

Further predictive performance measures that we shall consider do not depend on the choice of a cut off: the Receiver Operating Characteristic curve (ROC), the area under it (AUC) and the H measure (see e.g. Hand et al. 2010).

So far we have discussed the issue of predictive performance. Another important characteristic that a model should have is discriminant power. The *discriminant power* of a predictive model can be measured by a confusion matrix which compares actual and predicted classifications for a fixed cut-off. From the confusion matrix we can derive the percentage of correct classification, the sensitivity (the percentage of default events correctly identified by the model) and the specificity (the percentage of non default events correctly identified by the model).

3 Proposal

The choice of a prior distribution over the model space remains an open area of research in model averaging approaches. A common assumption for a prior on the model space is to assume that each model is equally likely, with all prior model probability equal to $\frac{1}{2^q}$, where 2^q is the total number of models to be considered. In this case the posterior model probabilities are determined only by the marginal likelihoods and therefore only the priors on the parameters affects the results of BMA.

Fernandez et al. (2001), Ley and Steel (2007) and Eicher et al. (2011) argue that prior assumptions can be extremely critical for the outcome of the analyses, and that uniform priors are not neutral at all. Instead, they suggest, in the context of linear regression models, to use hierarchical priors, that lead to more stable results, in terms of posterior model probabilities. Here we extend this approach and, in particular, Ley and Steel (2007), to logistic regression models, widely employed in credit risk modelling starting from the seminal paper of Altman 1968. Hierarchical model priors can be assigned considering that a model can be seen as made up by "marking" each of the available variable as present or not in the model. It becomes thus natural to consider a model to be described by the number of variables to be included, with the latter being modelled as a Binomial distribution. According to this prior, each variable is independently included (or not) in a model so that the model size M follows a Binomial distribution with probability of success θ: $M \sim Bin(q, \theta)$ where q is the number of candidate regressors considered and θ is the prior inclusion probability for each variable.
Given the above, the prior probability of a model M_j, with m regressors, is given by:

$$P(M_j) = \theta^m (1-\theta)^{q-m}. \quad (3.1)$$

As a special case of this prior structure, if we assume that every model has the same a priori probability, we obtain a uniform prior on the model space, which is what assumed by many authors including Hoeting et al. (1999). This uniform prior corresponds to the assumption that $\theta = 0.5$ so that the previous equation reduces to: $P(M_j = 2^{-q})$.

Ley and Steel (2007) propose an alternative prior specification in which θ is treated as random rather than fixed. The proposed hierarchical prior implies a substantial increase in prior uncertainty about the model size M, and makes the choice of prior model probabilities much less critical.
In particular, their proposal is the following:

$$M \sim Bin(q, \theta), \tag{3.2}$$

$$\theta \sim Beta(a, b) \tag{3.3}$$

where $a, b > 0$ are hyper-parameters to be fixed by the analyst. If we choose a Beta prior for θ with hyper-parameters $a, b > 0$, the prior mean model size is $E(M) = (a/a + b) * q$. The implied prior distribution on the model size is then a Binomial-Beta distribution. In the case where $a = b = 1$ we obtain a discrete uniform prior for model size with $P(M = m) = 1/(q + 1)$. This prior depends on two parameters (a, b) and Ley and Steel (2007) propose to facilitate prior elicitation fixing $a = 1$. This still allows for a wide range of prior distributions and makes it attractive to elicit the prior in terms of the specification prior mean model size $E(M) = m$. The number of regressors, q will then determine b through $b = (q - m)/m$.
Thus, in this setting, the analyst only needs to specify a prior mean model size m, which is exactly the same information one needs to specify for the case with fixed θ which should than equal $\theta = m/q$.
We remark that both the Binomial and the Binomial-Beta priors have in common the implicit assumption that the probability of one regressor appears in the model is independent of the inclusion of others, whereas regressors are typically correlated. This is related to the dilution problem raised by George (1999).
On the basis of the previous remarks, our proposal for model averaging can be summarised in the following steps:

- Given q candidate regressors, we consider all the possible variables combination and we obtain the model space $\mathbf{M}$ of dimension 2^q.

- For each model we compute its marginal likelihood. The marginal likelihood in logistic regression analysis under model M_j can be obtained resorting to the approximation provided by Chib (1995) and Groenewald et al. (2005), which is the approach we follow here. An alternative approach would be that of Holmes and Held (2006) who adopt auxiliary variable approaches.

- We assume the Ley and Steel (2007) on the model space, with different specifications of the hyper parameters involved, and we compare them in a sensitivity analysis framework.

- For each model we obtain its posterior model probability.

- For each model we derive all predictive values for the events to be forecasted and then we calculate the final forecast for a specific event is the average of the predictions made by each model weighted by the relative posterior model probability.

4 Empirical Evidences

4.1 Musing data

Our empirical analysis is based on two different data bases in credit risk analysis. We first concentrate on a database that is based on annual 1996–2004 data from Creditreform, one of the major rating agencies for Small and Medium Enterprises (SME) in Germany, obtained from the European Musing Project. In this case we have a small number of available covariates selected in advance.

When handling bankruptcy data it is natural to label one of the categories as success (healthy) or failure (default) and to assign them the values 0 and 1 respectively. Our data set consists of a binary response variable (default) values Y_i and a set of 4 explanatory variables $X_1, \ldots, X_4$, selected by the company. More precisely we have the following candidate regressors:

- *Equity ratio =(V1)*: it measures the financial leverage of a company calculated by dividing a measure of equity by the total assets.

- *Liabilities ratio=(V2)*: it is a measure of financial exposure calculated by dividing a gross measure of long-term debt by the assets of the company.

- *Result ratio=(V3)*: this is an index of how profitable a company is relative to its total assets.

- *ZwsUrt=(V4)*: this variable summarises the payment history of each SME company. The levels are 0 if the payment is within time and 1 if irregular payments are present.

We have constructed all the possible model combinations and precisely: 16 models in total with different complexities. The resulting model space is reported in Table 1:

Table 1 about here

As we can observe from Table 1, for $K = 0$ the number of corresponding models is equal to 1; for $K = 1$ the number of possible models is equal to 4; if $K = 2$ the number of possible models is equal 6; if $K = 3$ the number of possible models is equal to 4 and if $K = 4$ the number of possible models is equal to 1 and it corresponds to the saturated model.
We now specify the prior distribution on the model space. Following Ley and Steel (2007) we employ a hierarchical approach assuming θ random. The results for each model in terms of posterior probabilities are reported in Table 2. In Table 2 we report besides posterior model probabilities also predictive performance measures and specifically the H and the AUC indexes.

Table 2 about here

From Table 2 we observe that M4 is the best model in terms of posterior probability, for all considered parameterisations of the Beta prior distribution. However, it is not the best in terms of predictive performance, as model M1 is clearly superior both in terms of AUC and H. However, the model chosen by a standard, single model, Bayesian analysis would be M4, because it shows the highest posterior model probability. Therefore, a standard Bayesian approach loses predictive performance. Note however that the model that would be selected by a classical logistic regression approach (CLR) is M11, which performs considerably worse than the standard Bayesian model.

A second remark from Table 2 is that the use of a hierarchical prior reduces the dependence of the posterior distribution (and, therefore, of predictive results) on prior assumptions: this can be seen observing that in Table 2, regardless of the chosen prior hyperparameters, model posteriors are quite stable. This would not be the case with a non-hierarchical prior, as our results show.

We now proceed with a Bayesian model averaged approach, whose predictive power will be assessed using the AUC and the H index, as before. It turns out that the averaged model using $\theta \sim Beta(1, 1)$ reports an AUC equal to 0.910 and an H index equal to 0.377 while the averaged models using $\theta \sim Beta(1,2)$ and $\theta \sim Beta(1,3)$ both report an AUC equal to 0.909 and an H index equal to 0.376. Similar results have been obtained on the same data by Figini and Giudici (2011), using a weighted average of two models one based only on quantitative variables and the other based only on qualitative variables.

In order to draw further comparisons among the models at hand in terms of discriminatory power we have derived, for two different cut-offs: 0.5 a "neutral value" and 0.125, which correspond to the observed frequency of defaults, the sensitivity, the specificity and the percentage of correct classifications. Of main interest in credit risk model is the sensitivity, which measures the capability of the model to correctly predict the rarest event, the credit default. The results are in Table 3, using a Beta (1,1) prior distribution.

Table 3 about here

Table 3 shows that the Bayesian averaged model is superior in terms of sensitivity, especially for the more realistic 0.125 cut-offs, while keeping a similar percentage of correct classifications. Different parametrisations for the prior Beta distribution do lead to similar results.

4.2 Munich data

We now consider a second credit risk database, with a larger amount of predictors. This implies that we will not be able to perform exact calculation of model posterior probabilities, as the model space is too large. Rather, we shall follow the BMA approach suggested by Hoeting et al (1999) which includes an Occam's razor approach to reduce the number of models to be compared. In such an approach, model priors are taken as uniform.

The dataset we consider is the Munich dataset (see e.g. Giudici and Figini, 2009), and comprises, besides the target variable, 20 possible explanatory predictors which have been transformed in a binary scale. The results for the best models in terms of posterior probability are reported in Table 4. In Table 4 we report, besides posterior model probabilities, also the model characterisation in terms of variables included for the best 5 models.

Table 4 about here

From Table 4 we observe that M1 and M2 are the best in terms of posterior probability. The chosen single model, in a standard Bayesian analysis, would be M1, because it shows the highest posterior model probability. This should be compared with the model selected by a classical logistic regression which is not included in the best five models summarised in Table 4. Such a model selects as relevant the following variables: Deadline, Previous response, Purpose, Bank book, Working years, Monthly interests, Age and House.
We are now able to compare the predictive powers of the chosen Bayesian model (M1), the Bayesian Model average and the Classical Logistic Regression. It turns out that the averaged model using BMA reports an AUC equal to 0.77 and an H index equal to 0.09 while the chosen Bayesian model report an AUC equal to 0.69 and an H index equal to 0.065. Finally, the AUC for the classical logistic regression is equal to 0.77 and the H index equal to 0.09. Therefore, for the larger database, the predictive performance of BMA is similar to that of the selected logistic regression model, with the best chosen Bayesian model legging behind.
The good predictive performance of classical models with respect to Bayesian ones, on large data sets, is known in the literature. Our contributions shows that a model averaged Bayesian model can overcome this problem. In order to make a more precise comparison among the models we have derived for two different cut-offs: the neutral (0.5) and the observed frequency one (0.3), the sensitivity, the specificity and the percentage of correct classifications. The results are in Table 5.

Table 5 about here

From Table 5 note that the best model is the Bayesian model averaged, for both cut-offs in terms of sensitivity. It is slightly superior in terms of sensitivity, with respect to the CLR, with a slightly better performance also in terms of percentage of correct classifications. The single best Bayesian model clearly underperforms.

5 Conclusions

In this paper we have proposed a novel approach for credit risk modelling, that use Bayesian model averaging. Our aim is to obtain a good model that predicts well credit default events, on the basis of the estimated probability of default. In the paper we suggest the use of hierarchical priors, that lead to

more stable results in terms of posterior model probabilities and, therefore, in terms of predictive performances.
A non uniform prior over the model space is suitable when the number of covariates at hand is small (as in our first application), so that specific subject matter considerations can be employed to assess prior opinions over the model space. When the number of variables is large (as in our second application), instead we suggest to consider a model averaging approach with a uniform prior over the model space.
In both applications the proposed model averaging approach overperforms classical logistic regression models and best single Bayesian models in terms of accurate predictions of default events, while keeping the percentage of correct classifications at roughly the same level.
We can suggest our approach to credit risk default prediction especially when the main aim of the analysis is a selective detection of default events, rather than a detection of both good and bad occurrences, for which a simpler classical logistic regression model may be more appropriate.

Table 1: Model Space for the Musing data

Model	V1	V2	V3	V4
M1	1	1	1	1
M2	1	1	1	0
M3	1	0	1	1
M4	0	1	1	1
M5	1	1	0	1
M6	1	1	0	0
M7	1	0	1	0
M8	0	0	1	1
M9	0	1	0	1
M10	1	0	0	1
M11	0	1	1	0
M12	1	0	0	0
M13	0	1	0	0
M14	0	0	1	0
M15	0	0	0	1
M16	0	0	0	0

6 Tables

Table 2: Model posterior probabilities for the Musing data

Modelli	K	$\theta \sim Beta(1,1)$	$\theta \sim Beta(1,2)$	$\theta \sim Beta(1,3)$	H	AUC
M1	4	0.37	0.22	0.16	0.379	0.911
M4	3	0.50	0.61	0.66	0.374	0.908
M3	3	0.14	0.17	0.18	0.374	0.907
M8	2	0.00	0.00	0.00	0.274	0.879
M5	3	0.00	0.00	0.00	0.337	0.877
M10	2	0.00	0.00	0.00	0.332	0.8746
M9	2	0.00	0.00	0.00	0.334	0.874
M2	3	0.00	0.00	0.00	0.161	0.843
M7	2	0.00	0.00	0.00	0.154	0.838
M11=CLR	2	0.00	0.00	0.00	0.157	0.836
M15	1	0.00	0.00	0.00	0.14	0.786
M6	2	0.00	0.00	0.00	0.097	0.775
M14	1	0.00	0.00	0.00	0.116	0.769
M13	1	0.00	0.00	0.00	0.089	0.768
M12	1	0.00	0.00	0.00	0.085	0.766
M16	0	0.00	0.00	0.00	0.12	0.686

Table 3: Predictive model assessment for the Musing data

Chosen Bayesian Model (M4)	P=0.5	P=0.125
cut-off	0.5	0.125
sensitivity	0.543	0.4259
specificity	0.958	0.9662
% correct	0.906	0.8405
Bayesian Model Average	P=0.5	P=0.125
cut-off	0.5	0.125
sensitivity	0.543	0.4325
specificity	0.959	0.8017
% correct	0.907	0.843
CLR Model	P=0.5	P=0.125
cut-off	0.5	0.125
sensitivity	0.05	0.43
specificity	0.985	0.9691
% correct	0.868	0.844

Table 4: Selected Models (BMA)

Variable	M1	M2	M3	M4	M5
Intercept	1	1	1	1	1
Deadline	1	1	1	1	1
Previous Rep	1	1	1	1	1
Purpose	1	1	1	1	1
Bank book	1	1	1	1	1
Working years	1	1	1		1
Monthly interests		1			1
Age	1	1		1	1
House	1	1	1	1	1
Foreign	1	1	1	1	
Posterior Model Probability	0.127	0.123	0.09	0.087	0.083

Table 5: Predictive model assessment for the Munich data

Chosen Bayesian Model (M1)	P=0.5	P=0.3
cut-off	0.5	0.3
sensitivity	0.08	0.03
specificity	0.98	1
% correct	0.714	0.711
BMA	P=0.5	P=0.3
cut-off	0.5	0.3
sensitivity	0.78	0.72
specificity	0.65	0.85
% correct	0.75	0.73
CLR Model	P=0.5	P=0.3
cut-off	0.5	0.3
sensitivity	0.74	0.71
specificity	0.8	0.85
% correct	0.75	0.71

References

[1] Bernardo and A. Smith (1994): Bayesian Theory, Wiley, New York.

[2] Chib S. (1995): Marginal likelihood from the Gibbs output, Journal of the American Statistical Association, 90, 1313-1321

[3] Claeskens, G. and Hjort N. (2003): The Focused Information Criterion, Journal of the American Statistical Association, 98, 900-916.

[4] Durlauf, S., A. Kourtellos, and C. Tan(2008): Are Any Growth Theories Robust? Economic Journal, 118, 329-346.

[5] Eicher,T.,C. Henn, and C.Papageorgiou (2009a): Trade Creation and Diversion Revisited: Accounting for Model Uncertainty and Natural Trading Partner Effects, Journal of Applied Econometrics, forthcoming.

[6] Eicher, T., A. Lenkoski, andA. Raftery (2009b): Bayesian Model Averaging and Endogeneity Under Model Uncertainty: An Application to Development Determinants, University of Washington Working Paper UWEC-2009-19.

[7] Eicher, T., C. Papageorgiou, and A. Raftery (2011): Default Priors and Predictive Performance in Bayesian Model Averaging, with Application to Growth Determinants, Journal of Applied Econometrics, forthcoming.

[8] Fernandez, C., E. Ley, and M. Steel (2001): Benchmark Priors for Bayesian Model Averaging, Journal of Econometrics, 100, 381-427.

[9] Figini S. and Giudici P. (2011): Statistical Merging of Rating Models, Journal of the Operational Research Society, vol. 62, 1067-1074.

[10] George, E. (1999): Discussion of Model Averaging and Model Search Strategies by M. Clyde, in Bayesian Statistics, ed. by J. Bernardo, A. Berger, P. Dawid, Oxford University Press.

[11] Giudici P. and Figini S. (2009): Applied Data Mining, Wiley.

[12] Groenewald P.C.N. and Mokgatlhe L. (2005): Bayesian Computation for logistic regression, Computational Statistics and Data Analysis, 48, 857-868.

[13] Kass, R. and L. Wasserman(1995): A Reference Bayesian Test for Nested Hypothesis with Large Samples, Journal of the American Statistical Association, 90, 928-934.

[14] Hand, D. and Krzanowski W.J. (2010): ROC curves for continuous data. CRC/Chapman and Hall.

[15] Hoeting, J.A., Madigan, D., Raftery, A.E. and Volinsky, C.T. (1999). Bayesian model averaging: A tutorial (with Discussion). Statistical Science, 14, 382-401.

[16] Hayden, E., Stomper, A. and Westerkamp A. (2009). Selection vs Averaging of Logistic Credit Risk models, submitted.

[17] Holmes, C. and Held, K. (2006): Bayesian Auxilliary Variable Models for Binary and Multinomial Regression, Bayesian Analysis, 1, 1, 145-168.

[18] Ley, E. and M. Steel (2007): Jointness in Bayesian Variable Selection with Applications to Growth Regressions, Journal of Macroeconomics, 29, 476-493.

[19] Madigan, D. and A. Raftery (1994): Model Selection and Accounting for Model Uncertainty in Graphical Models using Occams Window, Journal of theAmerican Statistical Association, 89, 1535-1546.

[20] Madigan, D. and J. York (1995): Bayesian Graphical Models for Discrete Data, International Statistical Review, 63, 215-232.

[21] Schroder, B. and Richter, O. (1999). Are habitat models transferable in space and time? Journal of Nature Conservation, 8, 195-205.

[22] Steel M.F.J (2011): Bayesian Model Averaging and Forecasting, Working paper.

[23] Tsai C.L., Wang H. and Zhu N. (2010). Does a Bayesian approach generate robust forecasts? Evidence from applications in portfolio investment decisions, *Annals of the Institute of Statistical Mathematics*, vol. 62, pp. 109-116.

[24] Zweig, M.H. and Campbell, G. (1993). Receiver-Operating Characteristic (ROC) Plots: a fundamental evaluation tool in clinical medicine, Clinical chemistry 39, 561- 577.