Bayesian model averaging for financial evaluation

Silvia Figini

Abstract In this paper we propose classical and Bayesian Model Averaging (BMA) models for logistic regression in credit risk measurement. On the basis of a real data set, we show that Bayesian Model Averaging models outperform classical logistic regression in terms of percentage of correct classifications and related performance indicators.

Key words: Model selection, Bayesian Model Averaging, Credit Risk

1 Introduction

In credit risk 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. Relying upon a single model may not be the best strategy. In this paper we investigate whether the usage of more models, in a model averaging perspective, improves the performance of credit risk models. To achieve this aim we propose, starting from the most employed credit risk model - logistic regression - to average results obtained from a collection of models, using either bootstrapping or a Bayesian perspective.
In order to show how our proposal works, we have used a real data set provided by a credit rating agency composed of small and medium enterprises, that has already been considered in the statistical literature.
The paper is structured as follows: in Section 2 we review the Bayesian model averaging techniques; Section 3 reports empirical evidences achieved on a financial data set; Section 4 ends the paper and reports further ideas of research.

Silvia Figini
University of Pavia, e-mail: silvia.figini@unipv.it

2 Bayesian Model Averaging: a review

Typically, standard statistical practice ignores model uncertainty. BMA (Hoeting et al., 2001) provides a coherent mechanism for accounting for this model uncertainty. It is a technique designed to help account for the uncertainty inherent in the model selection process, something which traditional statistical analysis often neglects. By averaging over many different competing models, BMA incorporates model uncertainty into conclusions about parameters and predictions. BMA has been applied successfully to many statistical model classes including linear regression, generalized linear models, proportional hazard models, and discrete graphical models, in all cases improving predictive performance.
Let $\mathbf{M} = (M_1, ..., M_k)$ be the set of models under consideration. A model may be defined by a variety of attributes such as the subset of explanatory variables in the model or the form of the error variance. If Δ is the quantity of interest, such as a future observable or a model parameter, then the posterior distribution of Δ given data Z is equal to:

$$p(\Delta|Z) = \sum_{k=1}^{K} p(\Delta|Z, M_k) p(M_k|Z). \tag{1}$$

This is an average of the posterior predictive distribution under each of the models considered, weighted by the corresponding posterior model probability. The posterior probability for model M_k is given by:

$$p(M_k|Z) = \frac{p(Z|M_k)p(M_k)}{\sum_{l=1}^{K} p(Z|M_l)p(M_l)},$$

where $p(Z|M_k)$ is the integrated likelihood of model M_k, θ_k is the vector of parameters of model M_k, $p(\theta_k|M_k)$ is the prior density of the parameters under model M_k, $p(Z|\theta_k, M_k)$ is the likelihood and $p(M_k)$ is the prior probability that M_k is the true model. All probabilities are implicitly conditional on $\mathbf{M}$, the set of all models being considered. A number of researchers have considered the problem of managing the summation in equation (1) for a large number of models, some of which are described below for specific areas of application. A popular approach is to explore the space of models stochastically via a Markov chain Monte Carlo approach.
Hoeting, Madigan, Raftery, Volinsky (2001) discuss the historical development of BMA, provide additional description of the challenges of carrying out BMA, and describe some solutions to these problems for a variety of model classes.
In this paper we focus on BMA for generalised linear models with a specific focus on predictive models for binary outcome.

3 Application

Our empirical analysis is based on annual 1996–2004 data from Creditreform, one of the major rating agencies for SMEs in Germany.

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. Therefore, our data set consists of a binary response variable (default) values Y_i and a set of explanatory variables: $X_{1i},...,X_{ki}$ that are quantitative financial ratios and $X^*_{1i},...,X^*_{pi}$ that are qualitative features. The sample size available is composed of 1000 SMEs. The observed probability of default is equal to 12.5%.

In a one step perspective, we have used the whole data set, putting together quantitative and qualitative variables. Classical logistic regression analysis (CLR) allows to select the best predictive model and to obtain parameter estimates conditionally on such model. In order to assess predictive ability, we have implemented a cross-validation procedure: we have used 70% of the observations as a training set and 30% as a validation set on which to calculate predictive accuracy.

As discussed in the previous section, we have then tried to improve CLR analysis by means of a BMA approach. However, in order to better compare in terms of efficiency BMA with CLR we have carried out a bootstrapped version of BMA and CLR. The BMA analysis has been conducted assuming that all models are equally likely a priori and implementing the algorithm described in Hoeting et al., 2001.

We first check if the different models are different in terms of predictive ability. To reach this objective, on the basis of the percentage of correct classifications (PC), we have compared the models. Table 1 reports the results obtained using a large set of cut offs. From Table 1 we note that fixing a cut off greater than 0.4 the BBMA is

cut off	CLR	BMA	BCLR	BBMA
0.1	0.8168	0.8125	0.125	0.8127
0.2	0.8728	0.8772	0.8545	0.857
0.3	0.8987	0.8944	0.9084	0.899
0.4	0.9073	0.9019	0.8922	0.912
0.5	0.9106	0.9084	0.875	0.912
0.6	0.9084	0.8998	0.875	0.912
0.7	0.8901	0.8847	0.875	0.9234
0.8	0.875	0.8761	0.875	0.9238
0.9	0.8739	0.8739	0.875	0.938
1	0.875	0.875	0.875	0.938

Table 1 Correct classification for different cut offs

the best model in terms of correct classifications. We have then computed the AUC with the relative confidence interval on the basis of a bootstrap percentile method (see e.g. Hosmer et al, 2000). Table 2 shows for all models the AUC and the correct classification rate. From Table 2 the BBMA model shows the best performance. In addition we have obtained that the AUC test based on the confidence intervals confirms that the AUCs computed are significantly different from 0.7. In business

Model	AUC
Classical Logistic Regression	0.837
BMA Logistic Regression	0.8916
Boostrapped Classical Logistic Regression	0.913
Boostrapped BMA Logistic Regression	0.945

Table 2 Model assessment based on AUC

practice, this means that the discrimination made by the corresponding model is acceptable. For the sake of comparison, we have also compared our approach based on one step integration with the results described in Figini and Giudici 2011 which are based on a two step model. In the merged model proposed in Figini and Giudici 2011, the obtained AUC is equal to 0.909 and the percentage of correct classification using a cut off equal to 0.8 is equal to 0.915.
Therefore, our proposal leads to better results in terms of model performance.
On the basis of the results achieved we think that Bayesian averaged models lead better results in terms of predictive performance, while classical approaches provide an efficient and parsimonious method to select the most important variables.

4 Conclusions

In this paper we have presented a comparison between classical and model averaged models for credit risk estimation. Our results suggest that model averaged models could be considered as an alternative to classical logistic regression. In particular, we have found that, in comparison with classical logistic regression models, Bayesian Model Averaging provides risk models with superior discriminatory power and comparable predictive performance.

Acknowledgements The author thanks Prof. Paolo Giudici and Prof. Lijun Gao for useful suggestions and comments.

References

1. Hoeting J., Madigan D., Raftery A. and Volinsky C. (2001). Bayesian Model Averaging, *Statistical Science* **14**, 382-401.
2. Figini, S. and Giudici, P. (2011). Statistical Merging of rating models, *Journal of the Operational Research Society*, **62**, 1067 -1074.
3. Hosmer, D.W. and Lemeshow, S. (2000). Applied logistic regression, Wiley New York.