LONDON
metropolitan
university

Centre for International Capital Markets

Discussion Papers ISSN 1749-3412

A Note Comparing Support Vector Machines and Ordered Choice Models' Predictions of International Banks' Ratings

Tony Bellotti, Roman Matousek, Chris Stewart

No 2009-3

A Note Comparing Support Vector Machines and Ordered Choice Models' Predictions of International Banks' Ratings

Tony Bellotti,[1] Roman Matousek[2] and Chris Stewart[3]

February 2009

Abstract

We find that Support Vector Machines virtually always predict international bank ratings better than ordered choice models.

Keywords: International bank ratings, support vector machines, ordered choice models.

JEL Classification: C25, C52, G21.

[1] Business School, University of Edinburgh, William Robertson Building, 50 George Square, Edinburgh EH8 9JY. Email: Tony.Bellotti@ed.ac.uk.

[2] Centre for International Capital Markets, London Metropolitan Business School, London Metropolitan University, 84 Moorgate, London, EC2M 6SQ. Tel: 020-7320-1569. E-mail: r.matousek@londonmet.ac.uk.

[3] London Metropolitan Business School, London Metropolitan University, 84 Moorgate, London, EC2M 6SQ. Tel: 020-7320-1651. E-mail: c.stewart@londonmet.ac.uk.

1. Introduction

Ratings of sovereign risk, corporate bonds and financial institutions conducted by rating agencies (RAs) may be seen as instruments that provide investors with *prima facie* information about the financial position of the subject in question and on the price of credit risk. Pinto (2006) argues that RAs opinions facilitate capital allocation through supplied information about the financial position of the companies in question. Indeed, the RAs' exclusive position may be justified because they reduce asymmetric information between investors and companies.

Ratings are ordinal measures that should not only reflect the current financial position of sovereign nations, firms, banks, etc. but also provide information about their future financial positions. There has been extensive research in predicting bond ratings using multi-variate discriminant analysis, ordered choice models, non-parametric techniques and combined methods' forecasts to predict *bond* ratings – see, Altman and Saunders, (1998), Kamstra *et al* (2001) and Kim (2005). Thus, we employ financial variables, in addition to country risk (which we model using country specific dummy variables), as determinants of *bank* ratings in our modelling. The main challenge in modelling ratings is to increase the probability of correct classifications. This motivates our comparison of Support Vector Machines (SVMs) with ordered choice models for predicting individual bank ratings as produced by Fitch Ratings (FR).[4]

[4] We consider SVMs rather than neural networks (NN) for the following reasons. SVM has good theoretical foundations in statistical learning theory (NN was not developed in this way). SVM find single global minima whereas NN may find local rather than global minima. NN have a greater tendency to overfit data than SVM due to the latter's complexity minimization property. The computational complexity of SVM does not increase with the size of input space as it does with NN.

The next section describes the data and methods applied while Section 3 discusses the principal empirical findings. The last section concludes.

2. Data and Methodology

FR is one of the largest rating companies for the banking industry around the world and releases four types of ratings: legal ratings, long term and short-term (security) ratings and individual ratings. We focus on individual ratings that assess the financial position of a bank itself. As stated by FR the rating is closely linked with financial performance (financial ratios). The individual rating provided by FR is divided into five broad categories according to the performances of rated banks and subdivided in to a total of nine categories.

We estimate models of international banks' ratings between 2000 and 2007.[5] This variable is ordinal and has up to nine ranked categories that are assigned integer values from 1 to 9 (in brackets) thus, E (1), D/E (2), D (3), C/D (4), C (5), B/C (6), B (7), A/B (8), A (9) – lower values indicate lower ratings. However, because four lagged values of covariates are used in our models all banks in category A are excluded from the sample leaving categories 1 to 8 (the sample size is 359/360 observations).

We use two methods for predicting bank ratings. The first is the ordered choice model that is well known to be appropriate for modelling an ordinal dependent variable (see Greene 2008) and is the standard method for modelling bank ratings. The second is

[5] The BankScope database has been used to obtain a large sample of commercial banks rated by FR.

the SVM. SVM is an approach that allows us to use complex non-linear models such as polynomial or Gaussian models efficiently using kernels. Unlike standard methods, such as logistic regression, SVM is designed to deal with a large number of covariates (features) since it includes a regularization term to control model complexity whilst fitting the data. Not only is this essential when fitting complex non-linear models but it proves an advantage for the bank rating problem when there are a large number of covariates available such as multiple country dummy variables (codes).

Since the SVM classifier is binary it is not immediately suitable for the bank rating problem. Multiple bank ratings can be modelled using multiple SVM classifiers in a "one-against-one" or directed acyclic graph approach (Huang *et al* 2004). However this requires many SVMs, which further increases model complexity and computation time. A simpler method is to use SVM regression to model ratings directly. An advantage that SVM regression has in contrast to classical regression techniques such as OLS is that the loss function is ε-insensitive meaning that differences between the predicted and true value less than ε are not treated as errors. Since we are interested in predicting integers, predictions are rounded to the nearest integer and so are indifferent to the fractional part of the prediction; for example, if the target rating is 2, then predictions of 2.1 or 2.3 are equally valid. Hence we can set $\varepsilon \leq 0.5$ to represent this indifference. This property allows SVM to be a more sensitive model for bank ratings.

SVM regression is expressed formally as follows. Given n observations (x_i, y_i) where x_i is a vector of covariates and y_i is a real number outcome, then SVM regression constructs a linear model $\hat{y} = w \cdot x + b$ by solving the quadratic optimization problem:

$$\min_{w,\xi,\xi^*}\left(\frac{1}{2}w\cdot w + C\sum_{i=1}^{n}\xi_i + \xi_i^*\right)$$

$$\text{subject to}\begin{cases} y_i - w\cdot x_i - b & \leq & \varepsilon + \xi_i \\ y_i - w\cdot x_i - b & \geq & \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* & \geq & 0 \end{cases}$$

This is essentially a least absolute value regression method with the ε-insensitive loss function:

$$L(y,\hat{y}) = \begin{cases} 0 & \text{if } |y-\hat{y}| \leq \varepsilon \\ |y-\hat{y}| - \varepsilon & \text{otherwise} \end{cases}$$

and a regularization term to minimize the magnitude (complexity) of *w*. The relative importance of the two optimization goals of fitting the data and reducing model complexity is controlled by the parameter *C*. Lower values give greater emphasis to reducing model complexity whilst larger values of *C* give greater emphasis to model fit. Increasing *C* should give better model fit on the training set (in-sample) but this will not always be translated into better fit on the test set (out-of-sample) since improvements may be due to over-fitting.

This representation is in primary form, but it can be transformed into dual form using Lagrange multipliers. In dual form non-linear models can be implemented efficiently using kernel methods. Typical kernels are polynomial and Gaussian kernels (Vapnik 1995). We used LIBSVM, a popular implementation of SVM by Chang and Lin (2001) at National Taiwan University.

We apply the SVM method to bank ratings data and compare its in-sample predictive performance with that of the current standard method for modelling ratings; ordered choice models. Various combinations of ε and C are considered in the SVM application.

We consider three sets of covariates to model bank ratings, being financial variables, the year in which the rating was made, [denoted $time_{it}$] and 89 country dummy variables (there are banks from 90 countries in total) to account for country-specific effects (capturing country risk). For the ordered choice models these country dummy variables could not all be entered simultaneously and so are combined in to a single index of indicators, following Hendry (2001).[6] Discussion of the exact method used to produce this index and specification of the index itself is given in Matousek and Stewart (2008). The 89 dummy variables were all entered together in the SVM application.

For the financial variables the first to fourth lagged values of the following are considered as potential determinants of bank ratings.[7] The ratio of equity to total assets [denoted $Equity_{it}$], the ratio of liquid assets to total assets [$Liquidity_{it}$] the natural logarithm of total assets [$\ln(Assets)_{it}$] and the net interest margin [*NI_Margin*]. Also considered are $NOA_{it} = OIA_{it} - OEA_{it}$ (where OIA_{it} is the ratio of operating income to total assets and OEA_{it} is the ratio of operating expenses to assets), the ratio of operating expenses to total operating income [$OEOI_{it}$] and the return on equity [$ROAE_{it}$].[8]

6 Hendry's analysis is within the context of modelling inflation using time-series data. Hendry and Santos (2005) discuss the potential advantages of using such an index.

7 Ordered choice models could not be estimated when the lag length exceeded four.

8 The following three further variables were also considered for inclusion in the model: the ratio of operating expenses to assets [OEA], the ratio of operating income to assets [OIA] and the return on

Although rating agencies would always endeavour to incorporate the most recent information into their ratings they typically form their views based on the history of a bank's performance. This justifies the consideration of variables lagged more than one period as covariates.

We do not include current values of these seven financial covariates because they may contain information that was unknown at the time the rating was made. For example, if a bank's rating was decided in January 2007 then the value of any explanatory factor measured over the whole of 2007 would be unknown when the rating was made.

3. Results

Two versions of ordered logit and probit models are used to predict bank ratings: a general model (including all lags of the variables) and a favoured parsimonious specification obtained using a cross-sectional variant of the general-to-specific methodology. These logit and probit models are reported in Table 1 and Table 3 of Matousek and Stewart (2008), respectively. The percentage of correctly predicted bank ratings from these ordered choice models are reported in Table 1. The percentages of correctly predicted bank ratings obtained from the SVM for various combinations of C and ε are reported in Table 2 under the section headed Unrestricted Dummies.[9]

---

assets [$ROAA$]. These were excluded from the models because they would cause a high degree of

The four ordered choice models have percentage correct predictions in the range of 50.6% to 51.5% which compares to the twenty-one SVM predictions of between 48.5% and 62.4% (the majority of SVM predictions exceed 57%). Thus, SVMs give substantially better in-sample predictive performance than ordered choice models. If such performance can be repeated out-of-sample this would suggest the adoption of SVMs would provide greater predictive accuracy than the methods currently used as standard for this purpose. This is important given that prediction is the primary purpose of such models.[10]

The in-sample predictive performance of SVMs is better for $\varepsilon < 0.5$ and C > 0.5 (relative to ordered choice models) and is best when $\varepsilon = 0.25$ and $C = 2$ (with 62.4% correct predictions). This suggests that choice of these parameters is important in the selection of the SVM used for prediction. Using an in-sample (or *ex-post* forecast) data set to identify the appropriate SVM for out-of-sample (*ex-ante*) prediction would seem to be a fitting strategy to adopt. Examining the correspondence of in-sample and out-of-sample predictive performance of SVMs relative to ordered choice models is a topic for further research that the authors intend to investigate.[11]

The SVM is re-estimated using the single country index to capture country effects instead of the 89 individual dummies and the results are reported in Table 2 (section

multicollinearity and their effects could be captured in other ways.

[9] SVM estimation results are not reported to save space but are available from the authors upon request.

[10] We deliberately did not use non-linear kernels to model the data in this exercise since we did not have sufficient data for an independent validation set to optimize kernel parameter settings. However we expect that using different kernels may improve performance and for future work we will apply SVMs with more complex non-linear models.

[11] Splitting the data set was not deemed appropriate in the current exercise because it would have restricted the sample to less than the 360 observations currently used for estimation. This is especially so given the difficulty already experienced in estimating ordered choice models with a large number of variables (for example, all the country dummies could not be included together in ordered choice models).

headed Single Country Index). The percentage of correct predictions for SVMs using the single country index are substantially lower than when the country dummies are unrestricted but are similar to those produced by the ordered choice models (that also use this country index).[12] Hence, it would seem that the superior predictive performance of SVMs over ordered choice models is because SVMs can be estimated including the large number of country dummies unrestrictedly, whereas the ordered choice models cannot.

4. Conclusions

We have found that SVMs can produce substantially better in-sample predictions of international bank ratings than the standard method currently used for this purpose, ordered choice models. This appears due to the SVM's ability to estimate a large number of country dummies unrestrictedly, which was not possible with the ordered choice models due to the sample size. Given that the primary purpose of modelling ratings is prediction this is an important result. Consideration of the relative out-of-sample predictive performance of SVMs and ordered choice models, requiring more observations than were available here, would be a desirable avenue for further research.

[12] The percentage correct predictions of the SVM with $\varepsilon = 0.25$ and $C = 4$ when country dummy variables are excluded is 32.2% which is a substantially worse performance relative to when country dummies are included. This confirms the finding of Matousek and Stewart (2008) concerning the importance of country dummies in predicting international bank ratings.

References

Altman, E. I., and Saunders, A. (1998). Credit risk measurement: Developments over the last 20 years. *Journal of Banking and Finance*, *21*, 1721–1742.

Chang C-C and Lin C-J (2001) LIBSVM : a library for support vector machines. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

Greene W. H., (2008). *Econometric Analysis*, Pearson, Prentice Hall, 6th edition.

Hendry D. F., (2001). Modelling UK Inflation, 1875 – 1991. *Journal of Applied Econometrics*, 16, 255 – 275.

Hendry D. F. and Santos C. (2005). Regression models with data-based indicator variables. *Oxford Bulletin of Economics and statistics*, 67, 5, 571 – 595.

Huang Z, Chen H, Hsu C-J, Chen W-H, Wu S (2004). Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision Support Systems* 37, 543-558.

Kamstra, M., Kennedy, P., Suan, T.K. (2001). Combining bond rating forecasts using logit. *The Financial Review*, 37, 75-96.

Kim, S. K. (2005). Predicting bond ratings using publicly available information, *Expert Systems with Applications*, 29, 75-81

Matousek and Stewart (2008). Ordered Choice Models of International Banks' Ratings with an Indicator Variable for Country Effects, Centre for International Capital Markets Discussion Paper 08-14, London Metropolitan University.

Pinto, A. R. (2006). Control and Responsibility of Credit Rating Agencies in the United States, *American Journal of Comparative Law*, 54, 341-356.

Vapnik V (1995). *The Nature of Statistical Learning Theory*. Springer NY

Table 1: Percentage Correct Predictions: Ordered Choice Models

	Logit	Probit
General	**51.5%**	50.1%
Favoured	50.6%	50.8%

The predicted rating for each observation is chosen upon the basis of the category with the highest probability.

Table 2: Percentage Correct Predictions: SVMs

		Unrestricted Dummies				**Single Country Index**		
		ε				ε		
		0.05	*0.25*	*0.5*		*0.05*	*0.25*	*0.5*
C	*0.25*	53.2%	53.8%	48.5%		52.9%	52.6%	49.6%
	0.5	56.8%	59.3%	51.0%		**53.2%**	52.4%	49.3%
	1	57.9%	61.3%	53.2%		52.9%	52.9%	48.7%
	2	58.2%	**62.4%**	56.3%		52.6%	52.9%	50.1%
	4	59.9%	61.6%	53.2%		52.9%	52.9%	51.8%
	8	58.5%	61.3%	53.2%		52.9%	52.1%	50.7%
	12	58.2%	61.6%	52.6%		52.6%	52.1%	51.3%