UNIVERSITY OF Southampton
School of Management

Performance measures of models to predict Loss Given Default: a critical review

Katarzyna Bijak

34th International Symposium on Forecasting
30th June 2014, Rotterdam

UNIVERSITY OF Southampton
School of Management

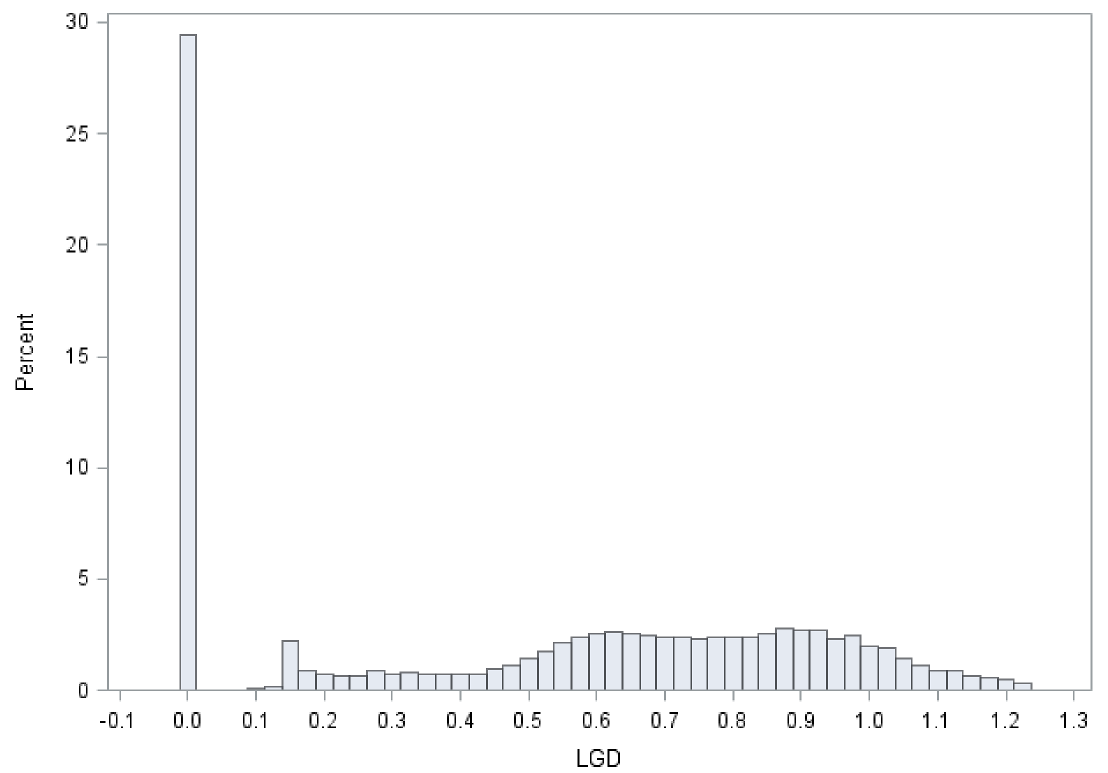
Outline

- Introduction
- Loss Given Default (LGD)
- Models to predict LGD
- Performance measures
  - Error measures
  - Other measures
- Conclusions

Loss Given Default (LGD)

- The lender's loss on a loan due to the customer's default, i.e. failure to meet the credit commitment
- "The ratio of the loss on an exposure due to the default of a counterparty to the amount outstanding at default" (Article 4(27) of the Council Directive 2006/48/EC)
- Basel II and III
  - Under the Advanced Internal Ratings-Based (AIRB) approach, lenders are allowed to use their own predictions of risk parameters, including LGD

LGD distribution example



Models to predict LGD

- Unsecured loans
  - One-stage models
  - Multi-stage approaches
    - Separation of 0s (+ Separation of 1s) + Prediction
- Mortgage loans
  - One-stage models
  - Two-stage approaches
    - Repossession model + Haircut model

Models to predict LGD

- Separation stage(s)
  - Logistic regression
  - Decision trees
- Prediction stage/One-stage models
  - Regression models
  - Tobit models
  - Survival analysis
  - Classification and Regression Trees (CART)
  - Other nonlinear models

Performance measures

- Credit scoring
  - Gini coefficient
  - Kolmogorov-Smirnov (KS) statistic
- LGD
  - ???

Error measures: MSE

- Mean Square Error (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- Sensitive to extreme values of the residuals
- E.g. Bellotti and Crook (2008)

Error measures: RMSE

- Root Mean Square Error (RMSE):

$$RMSE = \sqrt{MSE}$$

- Expressed in the same units as LGD
- Bastos (2010)

Error measures: MAE

- Mean Absolute Error (MAE) a.k.a. Mean Absolute Deviation (MAD):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

- Expressed in the same units as LGD
- Compare with RMSE
- E.g. Bellotti and Crook (2008)

UNIVERSITY OF Southampton
School of Management

Error measures: RAE

- Relative Absolute Error (RAE):

$$RAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{\sum_{i=1}^{n} |y_i - \bar{y}_i|}$$

- Ratio of MAE of the model and MAE of a simple predictor
- E.g. Bastos (2010)

UNIVERSITY OF Southampton
School of Management

Error measures: AOC

- Regression Error Characteristic (REC) curve estimates the CDF of the squared or absolute residual

- Area Over the REC Curve (AOC) estimates the expected regression error (Bi and Bennett, 2003)

- If the REC curve is derived using the squared (absolute) residuals, then AOC $\rightarrow$ MSE (MAE) as the sample size $\rightarrow \infty$

UNIVERSITY OF Southampton
School of Management

Error measures: AOC

- Loterman *et al.* (2012) calculated both RMSE and AOC (based on the squared residuals)
  - LGD models: 24 various techniques and six datasets
  - Differences between AOC and the squared RMSE:
    - < 0.001 for five larger datasets
    - < 0.01 for the smallest dataset (test: ca. 1100 loans)
- **We recommend applying either AOC or MSE/MAE in order to avoid information redundancy**

University of Southampton School of Management

Other measures: R-squared

- Coefficient of determination (R-squared):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$$

- E.g. Loterman *et al.* (2012)
- In an OLS regression model with a constant term, R-squared can be interpreted as the proportion of variation in LGD that is explained by variation in the regressors
- **We only recommend using R-squared in OLS models**

UNIVERSITY OF Southampton
School of Management

Other measures: Adjusted R-squared

- Adjusted coefficient of determination (adjusted R-squared):

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k-1}$$

- Corrected for the number of regressors (k)
- Useful when comparing a number of linear LGD models
- E.g. Caselli *et al.* (2008)

Other measures: Correlation coefficients

- Measure correlation between the observed and predicted LGD (Loterman *et al.*, 2012)

- Pearson's correlation coefficient:

$$r = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}}$$

  - Measures the strength of the linear relationship between the observed and predicted LGD ($r^2 = R^2$ in OLS models)

- Spearman's and Kendall's correlation coefficients

Other measures: AUC

- Loans need to be classified into two groups based on the observed LGD, e.g. below-the-mean and over-the-mean
- CDFs of the predicted LGD are computed for the groups
- Receiver Operating Characteristic (ROC) curve is drawn by plotting the CDFs against each other
- Area Under the ROC Curve (AUC) measures how well the model separates loans belonging to the two groups
- E.g. Gupton and Stein (2005)

Proposed measure: MAUC

- AUC has a drawback when applied to LGD as it requires an arbitrary classification of the dependent variable

- m – the number of unique values of the observed LGD

- Mean AUC (MAUC) is calculated as the average of AUC for all possible divisions into two groups:

$$MAUC = \frac{1}{m-1} \sum_{j=1}^{m-1} AUC_j$$

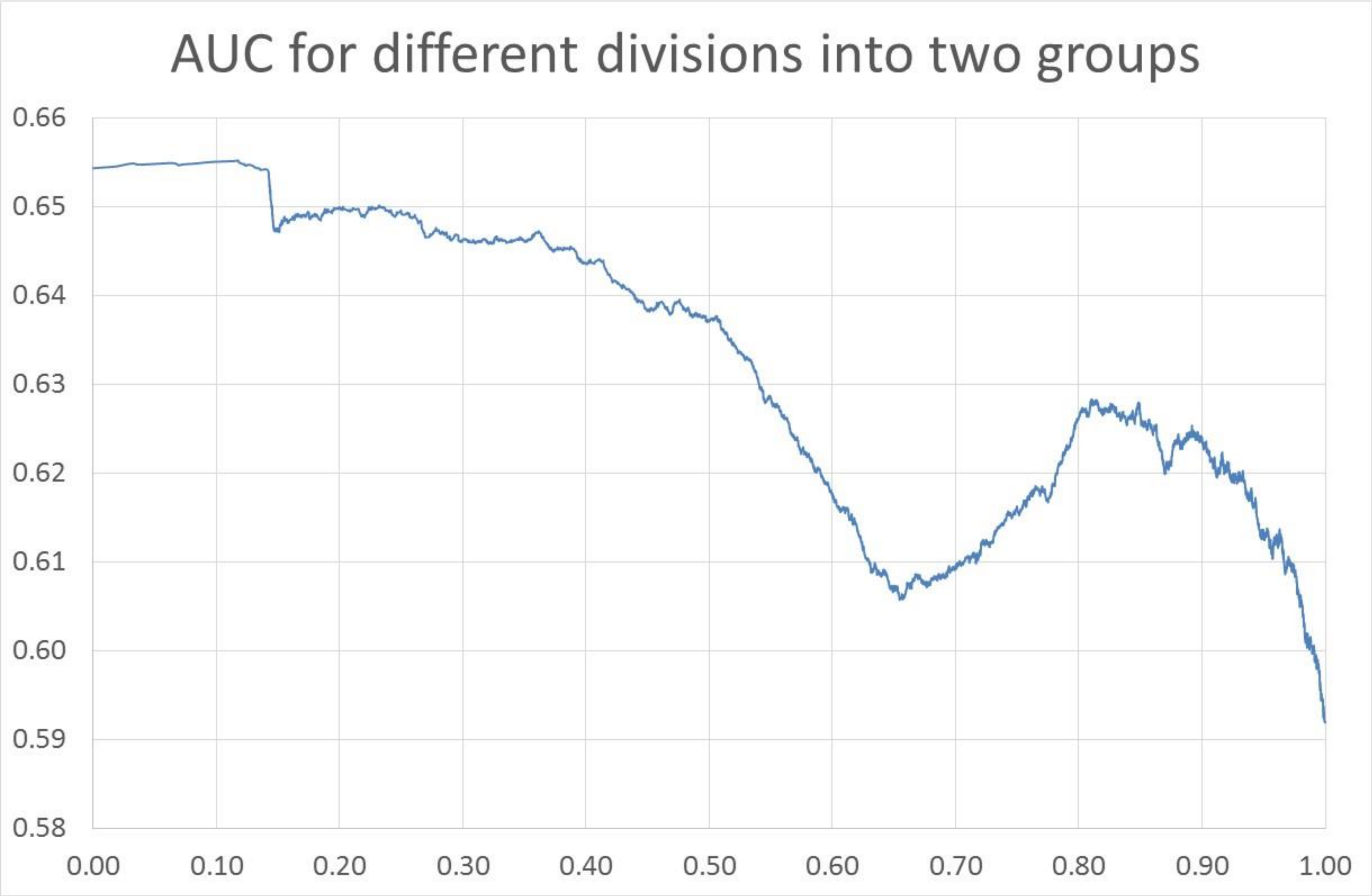
- MAUC takes values from the interval [0.5, 1] like AUC

Example

- Two-stage model applied to the data on personal loans granted by a large UK bank

Measure	Value	Measure	Value
MSE	0.143	Spearman	0.255
MAE	0.329	Kendall	0.179
R-squared	*0.072*	*AUC*	*0.637*
Pearson	0.268	MAUC	0.616

Example



AUC for different divisions into two groups
0.66
0.65
0.64
0.63
0.62
0.61
0.60
0.59
0.58
0.00
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.90
1.00

Conclusions

- Recommendations for LGD model developers/users
  - Apply either AOC or MSE/MAE
  - Only use R-squared in OLS models
  - Look for an alternative to AUC

- Further research
  - MAUC computed as the weighted average of AUC
  - Impact of segmentation on performance of LGD models

UNIVERSITY OF Southampton
School of Management

References

- Bastos, J.A. (2010) Forecasting bank loans loss-given-default, *Journal of Banking and Finance*, 34(10), pp. 2510-2517.
- Bellotti, T. and Crook, J. (2008) *Modelling and estimating Loss Given Default for credit cards*, University of Edinburgh Business School, Credit Research Centre Working Paper 08-1.
- Bi, J. and Bennett, K.P. (2003) Regression Error Characteristic Curves, In: Fawcett, T. and Mishra, N. (eds.) *Proceedings of the Twentieth International Conference on Machine Learning*, Menlo Park, CA: AAAI Press, pp. 43-50.
- Caselli, S., Gatti, S. and Querci, F. (2008) The Sensitivity of the Loss Given Default Rate to Systematic Risk: New Empirical Evidence on Bank Loans, *Journal of Financial Services Research*, 34(1), pp. 1-34.
- Gupton, G.M. and Stein, R.M. (2005) *LossCalc v2: Dynamic prediction of LGD*, Moody's KMV Research Paper.
- Loterman, G., Brown, I., Martens, D., Mues, C. and Baesens, B. (2012) Benchmarking regression algorithms for loss given default modeling, *International Journal of Forecasting*, 28(1), pp. 161-170.

UNIVERSITY OF Southampton
School of Management

Thank you!