Impact of segmentation on the performance measures of LGD models

Lyn Thomas and Katarzyna Bijak

Credit Scoring and Credit Control XIV
26th-28th August 2015, Edinburgh
Outline

• Introduction
• LGD
• LGD models
• Segmentation
• Example
• Conclusions
LGD

- **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

- **Recovery Rate (RR)**
  - \[ \text{LGD} = 1 - \text{RR} \]

- **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
LGD models

- 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
LGD models

• 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 of LGD models

- Please come on Friday!
  - Parallel session 22B - Performance Assessment
    - 12:05 - 12:35
Performance of LGD models

• Poor performance
  – In particular for unsecured consumer credit

• For example, Pearson’s correlation:
  – 0.38 - 0.51 (Yashkir and Yashkir, 2013)
  – 0.10 - 0.40 (Bellotti and Crook, 2008)
  – 0.08 - 0.27 (Bijak and Thomas, 2015)
Segmentation

• Poor performance should not be surprising
  – E.g. typical R-squared values when using regression for modelling outcomes over individual consumers are low

• Nevertheless, this is a bit embarrassing!

• ... but the Basel Accords only require LGD estimates at the portfolio segment level, where the lender has determined the appropriate segmentation of their portfolio
What happens at the segment level?

- Symbols
  - $n$ – number of individual observations
  - $K$ – number of segments
  - $M$ – segment size ($n = KM$)

- Assumptions for calculations
  - Linear regression model
  - Independent and normally distributed residuals with zero mean and the same variance
What happens at the segment level?

- Example: MAE

\[ \varepsilon_i \sim N(0, \sigma^2) \]

\[ MAE_n = \frac{1}{n} \sum_{i=1}^{n} |\varepsilon_i| \approx \sigma \sqrt{\frac{2}{\pi}} \]
What happens at the segment level?

- Example: MAE

\[ \sum \varepsilon_i \sim N(0, M\sigma^2) \]

\[
MAE_K = \frac{1}{K} \sum_{k=1}^{K} |\varepsilon_k| = \frac{1}{KM} \sum_{k=1}^{K} \left| \sum_{i=(k-1)M+1}^{kM} \varepsilon_i \right| \approx \frac{K\sigma}{KM} \sqrt{\frac{2M}{\pi}}
\]

\[ = \sqrt{\frac{K}{n}} MAE_n \]
What happens at the segment level?

- **MSE**

\[ MSE_K \approx \frac{K}{n} \text{MSE}_n \]

- **RMSE**

\[ RMSE_K \approx \sqrt{\frac{K}{n} \text{RMSE}_n} \]
What happens at the segment level?

• MAE

\[ MAE_K \approx \sqrt{\frac{K}{n}} \, MAE_n \]

• RSE

\[ RSE_K \approx 1 - \frac{n}{n - K + \frac{K}{1 - RSE_n}} \]
What happens at the segment level?

- R-squared

\[ R_K^2 \approx \frac{n}{n - K + \frac{K}{R_n^2}} \]

- Pearson’s correlation

\[ r_K \approx \sqrt{\frac{n}{n - K + \frac{K}{r_n^2}}} \]
Impact of segmentation

MSE

performance measure value

number of segments

10000  9000  8000  7000  6000  5000  4000  3000  2000  1000  0

0.16  0.14  0.12  0.10  0.08  0.06  0.04  0.02  0.00
Impact of segmentation (log. scale)
Impact of segmentation

![Graph showing the impact of segmentation on RMSE with decreasing RMSE as the number of segments increases.](image-url)
Impact of segmentation (log. scale)
Impact of segmentation
Impact of segmentation (log. scale)
Impact of segmentation
Impact of segmentation (log. scale)
Impact of segmentation

R-squared

number of segments

performance measure value
Impact of segmentation (log. scale)
Impact of segmentation

Pearson's correlation

Performance measure value

Number of segments
Impact of segmentation (log. scale)
Analytical vs empirical results

MSE

- Analytical
- Empirical

Performance measure value vs number of segments
Analytical vs empirical results

```
RMSE

- Analytical
- Empirical

performance measure value

number of segments
```

- Graph showing the performance measure value (RMSE) with respect to the number of segments for both analytical and empirical methods.
Analytical vs empirical results

MAE

- Analytical
- Empirical

- performance measure value

- number of segments
Analytical vs empirical results

The graph compares analytical and empirical results over the number of segments. As the number of segments increases, both analytical and empirical performance measure values decrease, indicating improved performance. The empirical results show a slightly steeper decline compared to the analytical results.
Analytical vs empirical results

R-squared

- Analytical
- Empirical

Performance measure value

Number of segments
Analytical vs empirical results
Conclusions

• Poor model performance at the loan level
• For a number of measures the segmentation leads to a significant improvement in their values
• Is this true for all performance measure?
• Do LGD models perform better than we thought?
Thank you!
Empirical results