A transformer-based model for default prediction in mid-cap corporate markets

KAMESH KORANGI, CHRISTOPHE MUES, CRISTIÁN BRAVO

CBRAVORO@UWO.CA
@CBRAVOR
The Banking Analytics Lab

One of the largest groups in the world researching analytics in banking, with a focus on credit risk.

- 3 postdoctoral researchers.
- 12 PhD students across three continents.
- A large network of collaborators all around the world.

Recognized worldwide!
- Markov et al. (2022):
  - Top 2 researcher in the area of credit scoring. Only in two lists (also in Louzada et al., 2017).
  - Two works (Bravo et al., 2013 and Verbraken et al., 2014) are now industry standards.

Visit our website!
https://thebal.ai
Agenda

- Mid-caps
- Problem
- Relevant Literature
- Contributions
- Data Description
- Results
- Conclusion
Mid-caps

Mid-cap companies are publicly traded companies with less than US$10 billion market capitalisation.

Their debt has a shorter maturity period than large-caps.

Typically hold a non-investment grade credit rating, implying higher credit risk.

Lots of disaggregated information, more dependant on general economic conditions.
Problem

Mid-cap challenges

• Credit spreads or prices implied by the models often underestimate from what is empirically observed\(^1\), and is more pronounced for mid-caps\(^2\)
• This mispricing of risk lead to unexpected losses for lenders
• Credit risk is not easily separable from market risk, especially for mid-cap companies\(^3\)

Modelling challenges

• Ability to use different kinds of data and extract relationships
• Time horizon of default prediction models\(^4\)
In the Literature...

Three types of credit risk models in this space – statistical models\(^5\), structural models\(^6\) and reduced form models\(^7\).

Machine learning models shown to improve performance, particularly ensemble of models in default prediction\(^8,9\).

Deep learning models have produced state-of-the-art results in other domains and been applied with textual data for corporates\(^10\) and for SMEs\(^11\) with promising results.

Our proposal

We construct an \textit{ad-hoc transformer model}\(^12\) -- state-of-the-art in natural language processing - for \textit{time series panel data with correlated outputs}. 
General Aim and Paper Contributions I

We develop a novel approach to credit risk modelling using deep learning in midcaps

- first to propose a transformer-encoder model for corporate default risk modelling
- a framework for multi-modal learning that can combine the different data sources and allows for a differential training approach

Single combined model for term structure of default probabilities

- Probabilities are obtained at 3m, 6m, 1y, 2y and 3y simultaneously.
General Aim and Paper Contributions II

We also make the model interpretable

- Utilize **attention heatmap visualizations** to show the model learning between defaults and survivals

- Shapley values to quantify the relative importance of **groups of variables** to answer
  - Which data sources are important
  - Which time period is important
Data description

• 30 years of data related to mid-cap companies listed in the US from 1990 to 2020

• Three different data sources

• Wider definition of default
Model architecture

Transformer Encoder model

Multi-modal architecture
Results

Deep learning models perform better with individual data

**Long-range dependencies** could be found with TransEncoder and TCN as can be seen from daily pricing data channel

With all data, **multi-modal architecture allows deep learning models to outperform XGB**
<table>
<thead>
<tr>
<th>Method</th>
<th>Regime</th>
<th>Average</th>
<th>d_3m</th>
<th>d_6m</th>
<th>d_9m</th>
<th>d_1y</th>
<th>d_2y</th>
<th>d_3y</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training together</td>
<td>2</td>
<td>0.812</td>
<td>0.829</td>
<td>0.821</td>
<td>0.813</td>
<td>0.808</td>
<td>0.802</td>
<td>0.799</td>
</tr>
<tr>
<td>Pricing channel freeze</td>
<td>3</td>
<td>0.817</td>
<td>0.828</td>
<td>0.833</td>
<td>0.822</td>
<td>0.812</td>
<td>0.805</td>
<td>0.802</td>
</tr>
<tr>
<td>Market and pricing channel freeze</td>
<td>3</td>
<td>0.821</td>
<td>0.839</td>
<td>0.814</td>
<td>0.820</td>
<td>0.814</td>
<td>0.826</td>
<td>0.812</td>
</tr>
<tr>
<td>TEP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training together</td>
<td>2</td>
<td>0.835</td>
<td>0.858</td>
<td>0.848</td>
<td>0.843</td>
<td>0.832</td>
<td>0.820</td>
<td>0.812</td>
</tr>
<tr>
<td>Pricing channel freeze</td>
<td>3</td>
<td>0.841</td>
<td>0.860</td>
<td>0.852</td>
<td>0.846</td>
<td>0.841</td>
<td>0.824</td>
<td>0.822</td>
</tr>
<tr>
<td>Market and pricing channel freeze</td>
<td>3</td>
<td>0.847</td>
<td>0.867</td>
<td>0.860</td>
<td>0.850</td>
<td>0.847</td>
<td>0.833</td>
<td>0.824</td>
</tr>
</tbody>
</table>
Interpreting the results

- Difference in learning between the default firms and non-default firms
- Each head focuses on different aspect of the input
- Input last 12 quarters of data mapped to output representation
- Higher weights colored yellow
Interpretation of the Results - Shapley

- Accounting information is most important data source for default prediction in mid-caps accounting for 30% importance
- Macro economic environment more important over medium term than equity performance of the company
- Pricing channel provides signalling in the short term
- Temporally, present data is more important for prediction over 52% of performance from present accounting information compared to 12.4% from past two years.

<table>
<thead>
<tr>
<th>Channel</th>
<th>Shapley values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Present 1 year</td>
</tr>
<tr>
<td>Fundamental</td>
<td>52.3</td>
</tr>
<tr>
<td>Market</td>
<td>35.1</td>
</tr>
<tr>
<td>Pricing</td>
<td>38.4</td>
</tr>
</tbody>
</table>
Conclusion

Deep learning models improve predictive power especially with complex models and multimodal architectures.

Term structure of probabilities produced within single model by manipulating the objective function.

Provide interpretability to the results by using heatmaps and custom methodology for groups of variables based on Shapley values.

Could extend models with unstructured data like text and audio data.

**Multi modal architecture** could be extended to produce new kind of scorecard models for credit risk.

Read the paper! Korangi et al. (2023) @ EJOR
Thank you!

Q&A
References


