Multiperiod Corporate Default Prediction Through Neural Parametric Family Learning

Wei-Lun Luo, Yu-Ming Lu, Jheng-Hong Yang, Jin-Chuan Duan, Chuan-Ju Wang
2022.4.28
Outline

• Introduction of Credit risk
  • Default risk
• Related works
  • Niche
• Methodology
• Results
Credit risk

- Lender: Banks
- Obligor: People, Companies

Lend money to obligors

Meet its obligations With the agreed terms
Credit risk

Lender
Banks

Meet its obligations
With the agreed terms

Obligor
People
Companies

Lend money to obligors
Credit risk

Lender
Banks

Obligor
People
Companies

Lend money to obligors
Meet its obligations with the agreed terms
Credit risk
Credit risk

Lender
Banks

Meet its obligations
With the agreed terms

Credit risk

Obligor
People
Companies

Lend money to obligors

Default risk
Default risk

A term structure of cumulative default probabilities (CDP)

Debts structure
- Short-term
- Long-term

Dissimilar with each other
Default risk

A term structure of cumulative default probabilities (CDP)

Debts structure
- Short-term
- Long-term

Dissimilar with each other

Monotonically Increasing
Default risk
A term structure of cumulative default probabilities (CDP)

Multiperiod default prediction
Multiperiod corporate default prediction
Related works

Default risk analysis

Risk classification (e.g. 3-months, 6 months)
Risk rankings


Related works

Default risk analysis

### Machine learning

**Risk classification** (e.g. 3-months, 6 months)

**Risk rankings**

**Real world applications**

E.g. multi-period


Related works
Default risk analysis

Machine learning
- Risk classification (e.g. 3-months, 6 months)
- Risk rankings

Statistical
- A consistent term structure of CDP
- Number of default occurrences

FIM

A term structure of CDP
Real world applications
E.g. multi-period

Related works

FIM

Performance deteriorates rapidly

CAP Curve

Related works

FIM

Related works

Default risk analysis

Machine learning
- Risk classification (e.g. 3-months, 6 months)
- Risk rankings

Statistical
- A term structure of CDP
- Number of default occurrences

FIM

- A consistent term structure
  Real world applications
  E.g. multi-period

- Rigorous assumption
  E.g. same parametric family
  for both long-term and short-term
Related works

Niche of each type of approaches

Machine learning

Risk classification (e.g. 3-months, 6 months)
Risk rankings

Rigorous assumption
E.g. same parametric family for both long-term and short-term

GAP

A consistent term structure
Real world applications
E.g. multi-period

Statistical

A term structure of CDP
Number of default occurrences

FIM
Related works
Leverage big data

Machine learning
Risk classification (e.g. 3-months, 6 months)
Risk rankings

Rigorous assumption
E.g. same parametric family for both long-term and short-term

GAP
A consistent term structure
Real world applications
E.g. multi-period

Statistical
A term structure of CDP
Number of default occurrences

FIM
Related works

Special model

Machine learning

Risk classification (e.g. 3-months, 6 months)

Risk rankings

GAP

A consistent term structure
Real world applications
E.g. multi-period

Design a special model

Statistical

A term structure of CDP
Number of default occurrences

Rigorous assumption
E.g. same parametric family for both long-term and short-term

FIM
Methodology
Neural Parametric Family Learning

Parameters

Mean
Standard Deviation

Parametric Family

Normal distribution
Pdf
Cdf

Parameters

- Mean
- Standard Deviation

Parametric Family

Normal distribution
Pdf
Cdf
Methodology

Two phase

\[ \mathcal{L} = \sum_{t \in T} \sum_{i=1}^{m} \sum_{k=1}^{m+1} \text{CrossEntropy}(\hat{F}_{t,i}(\tau_k), H_{t,i}(\tau_k)) \]

Objective Function

Cumulative Default Probability

Ground Truth Function

\[ \hat{F}_{t,i}(\tau_k) \]

\[ \tau_k \]

Parametric Family Parameter Generation \((\mathcal{N}_\theta)\)

\[ X_{t,i} = [x_{t-\delta+1,i}, \ldots, x_{t,i}] \in \mathbb{R}^{d \times \delta} \]

Input Covariate Matrix (\(\delta\) Lag Observations)

\[ \theta_{t,i} \in \mathbb{R}^p \]

Parameter

Parametric Family Determination \((\mathcal{N}_F)\)

Covariate Vector

\[ y_{t,i}[3+1] \]

\[ y_{t,i}[3] \]

\[ y_{t,i}[2] \]

\[ y_{t,i}[1] \]

MLP

Consider time-dynamics of covariates by time-lagged observations

Softmax

Add

Prediction Horizon
The first phase

The company set

The time point set

Methodology

A set of $\delta$ lag observations $X_{t,i}$

$d$-dimension vector for the $i$-th company at time point $t$ $X_{t,i}$

**Parametric Family Parameter Generation ($N_{\Theta}$)**

$$X_{t,i} = [x_{t-\delta+1,i}, \ldots, x_{t,i}] \in \mathbb{R}^{d \times \delta}$$

$$\theta_{t,i} \in \mathbb{R}^{p}$$

**Input Covariate Matrix ($\delta$ Lag Observations)**

**Parameter**

**RNN**

**Batch Normalization**

Consider time-dynamics of covariates by time-lagged observations

**Covariate Vector**

$x_{t-\delta+1,i}$ $x_{t,i}$
Methodology

The first phase

**Input**

- $X_{t,i}$: d-dimension vector for the i-th company at time point t
- $X_{t,i}^{2008/9,i} = [x_{2008/7,i}, x_{2008/8,i}, x_{2008/9,i}]$

**Parametric Family Parameter Generation** ($\mathcal{N}_0$)

\[ X_{t,i} = [x_{t-\delta+1,i}, \ldots, x_{t,i}] \in \mathbb{R}^{d \times \delta} \]

- Input Covariate Matrix ($\delta$ Lag Observations)
- Parameter $\theta_{t,i} \in \mathbb{R}^p$

**Consider time-dynamics of covariates by time-lagged observations**
The first phase

Methodology

The time point set $T$

The company set $I$

Input Covariate Matrix

(δ Lag Observations)

$X_{t,i} = [x_{t-\delta+1,i}, \ldots, x_{t,i}] \in \mathbb{R}^{d \times \delta}$

Parameter

$\theta_{t,i} \in \mathbb{R}^p$

Input

$d$-dimension vector for the $i$-th company at time point $t$

$X_{t,i}$

A set of $\delta$ lag observations $x_{t,i}$

Output

$p$-dimension vector for the $i$-th company at time point $t$

$\theta_{t,i}$

Consider time-dynamics of covariates by time-lagged observations
The second phase

The company can not survive forever

Marginal default probability

The time of the default

Methodology

Parametric Family Determination ($N_F$)

Objective Function

Ground Truth Function

Softmax

Add

Prediction Horizon

\[
\mathcal{L} = \sum_{t \in T} \sum_{i=1}^{n} \sum_{k=1}^{m+1} \text{CrossEntropy}(\hat{F}_{t,i}(\tau_k), H_{t,i}(\tau_k))
\]

\[
\hat{F}(\tau_\ell) = \sum_{k=1}^{\ell} y[k], \quad \text{for } \ell = 1, 2, \ldots, m + 1
\]

\[
H_{t,i}(s) = \begin{cases} 
1, & \text{if } s \geq \zeta_i \\
0, & \text{if } s < \zeta_i 
\end{cases}
\]
Methodology

The second phase

Parametric Family Determination \((N_F)\)

- \(y_{t,i}[3 + 1]\)
- \(\hat{F}_{t,i}(\tau_{3+1})\)
- \(y_{t,i}[3]\)
- \(\hat{F}_{t,i}(\tau_3)\)
- \(y_{t,i}[2]\)
- \(\hat{F}_{t,i}(\tau_2)\)
- \(y_{t,i}[1]\)
- \(\hat{F}_{t,i}(\tau_1)\)

Objective Function

\[ \mathcal{L} = \sum_{t \in T} \sum_{i=1}^{\mu} \sum_{k=1}^{m+1} \text{CrossEntropy}(\hat{F}_{t,i}(\tau_k), H_{t,i}(\tau_k)) \]

Cumulative Default Probability

\[ \hat{F}(\tau_\ell) = \sum_{k=1}^{\ell} y[k], \quad \text{for } \ell = 1, 2, \ldots, m+1 \]

Marginal default probability

The company can not survive forever

\[ H_{t,i}(s) = \begin{cases} 1, & \text{if } s \geq \zeta_i \\ 0, & \text{if } s < \zeta_i \end{cases} \]

The time of the default

Solve GAP

Real world applications

E.g. multi-period

A consistent term structure
Dataset
NUS Credit Research Initiative (CRI)

Dates: January 1990 - December 2017
Data: 1.5 M monthly samples of US public companies
Covariates: 14, 2 common and 10 firm-specific covariates
Events: 0 (alive), 1 (default), 2 (other exit)
Prediction horizons: 60 months
Experiment

Two types

Cross-sectional: randomly split data into 13 folds

<table>
<thead>
<tr>
<th>Cross-time</th>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1990 - 1999</td>
<td>2000</td>
</tr>
<tr>
<td>2</td>
<td>1991 - 2000</td>
<td>2001</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>2002 - 2011</td>
<td>2012</td>
</tr>
</tbody>
</table>
Metrics

Accuracy ratio (AR, %)

\[ AR = \frac{\text{Area above CAP curve}}{\text{Area under CAP curve}} \]

RMSNE

\[ \sqrt{\frac{1}{T} \sum_{i=1}^{T} \left( \frac{\hat{D}_i - D_i}{D_i} \right)^2} \]

\[ \hat{D}_i \]

\[ D_i \]

Estimated default occurrences (monthly)

Default occurrences (monthly)
### Results

Table 1: Results of cross-sectional experiments

<table>
<thead>
<tr>
<th>Horizons (months)</th>
<th>1</th>
<th>3</th>
<th>6</th>
<th>12</th>
<th>24</th>
<th>36</th>
<th>48</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy ratio (AR) (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIM</td>
<td>94.57</td>
<td>92.37</td>
<td>88.74</td>
<td>81.45</td>
<td>70.85</td>
<td>63.46</td>
<td>58.33</td>
<td>53.37</td>
</tr>
<tr>
<td>MLP (δ = 1)</td>
<td>94.48</td>
<td>92.85</td>
<td>90.43</td>
<td>85.10</td>
<td>75.63</td>
<td>68.08</td>
<td>62.87</td>
<td>58.26</td>
</tr>
<tr>
<td>MLP (δ = 6)</td>
<td>94.29</td>
<td>92.76</td>
<td>90.47</td>
<td>85.73</td>
<td>76.88</td>
<td>69.73</td>
<td>64.55</td>
<td>60.07</td>
</tr>
<tr>
<td>MLP (δ = 12)</td>
<td>93.99</td>
<td>92.64</td>
<td>90.55</td>
<td>86.05</td>
<td>77.67</td>
<td>70.81</td>
<td>65.93</td>
<td>61.45</td>
</tr>
<tr>
<td>LSTM (δ = 1)</td>
<td>94.78</td>
<td>93.17</td>
<td>90.87</td>
<td>86.11</td>
<td>77.47</td>
<td>70.69</td>
<td>65.70</td>
<td>61.09</td>
</tr>
<tr>
<td>LSTM (δ = 6)</td>
<td>94.63</td>
<td>93.29</td>
<td>91.23</td>
<td>87.05</td>
<td>79.00</td>
<td>72.63</td>
<td>67.55</td>
<td>62.96</td>
</tr>
<tr>
<td>LSTM (δ = 12)</td>
<td>94.68</td>
<td><strong>93.48</strong></td>
<td><strong>91.77</strong></td>
<td><strong>87.91</strong></td>
<td><strong>80.79</strong></td>
<td><strong>74.76</strong></td>
<td><strong>69.91</strong></td>
<td><strong>65.32</strong></td>
</tr>
<tr>
<td>GRU (δ = 1)</td>
<td>94.66</td>
<td>93.03</td>
<td>90.77</td>
<td>85.94</td>
<td>77.21</td>
<td>70.34</td>
<td>65.39</td>
<td>60.79</td>
</tr>
<tr>
<td>GRU (δ = 6)</td>
<td>94.41</td>
<td>92.97</td>
<td>90.84</td>
<td>86.54</td>
<td>78.26</td>
<td>71.60</td>
<td>66.45</td>
<td>61.91</td>
</tr>
<tr>
<td>GRU (δ = 12)</td>
<td>94.26</td>
<td>92.94</td>
<td>91.12</td>
<td>86.98</td>
<td>79.22</td>
<td>72.77</td>
<td>67.80</td>
<td>63.27</td>
</tr>
<tr>
<td>Improvement (%)</td>
<td>0.22</td>
<td>1.20</td>
<td>3.41</td>
<td>7.93</td>
<td>14.03</td>
<td>17.81</td>
<td>19.85</td>
<td>22.39</td>
</tr>
</tbody>
</table>

| Panel B           |     |     |     |     |     |     |     |     |
| Root mean square normalized error (RMSNE) |     |     |     |     |     |     |     |     |
| FIM               | 0.74 | 0.64 | 0.62 | 0.84 | 1.23 | 1.18 | 1.06 | 0.96 |
| MLP (δ = 1)       | 0.63 | 0.58 | 0.62 | 0.88 | 1.03 | 1.30 | 1.24 | 1.11 |
| MLP (δ = 6)       | 0.64 | 0.58 | 0.61 | 0.86 | 1.23 | 1.32 | 1.26 | 1.12 |
| MLP (δ = 12)      | 0.63 | **0.57** | **0.60** | 0.83 | 1.21 | 1.27 | 1.17 | 1.03 |
| LSTM (δ = 1)      | 0.62 | 0.60 | 0.64 | 0.89 | 1.26 | 1.30 | 1.23 | 1.11 |
| LSTM (δ = 6)      | 0.64 | 0.61 | 0.62 | 0.86 | 1.23 | 1.25 | 1.19 | 1.07 |
| LSTM (δ = 12)     | 0.64 | 0.62 | 0.61 | **0.81** | **1.11** | **1.12** | **1.03** | **0.90** |
| GRU (δ = 1)       | **0.61** | 0.61 | 0.65 | 0.91 | 1.25 | 1.32 | 1.23 | 1.11 |
| GRU (δ = 6)       | 0.64 | 0.63 | 0.64 | 0.87 | 1.24 | 1.29 | 1.22 | 1.11 |
| GRU (δ = 12)      | 0.64 | 0.64 | 0.64 | 0.83 | 1.13 | 1.18 | 1.10 | 0.98 |
| Improvement (%)    | 17.57 | 19.94 | 3.23 | 3.57 | 9.76 | 5.08 | 2.83 | 6.25 |
## Results

### Table 2: Results of cross-time experiments

<table>
<thead>
<tr>
<th>Horizons (months)</th>
<th>Panel A</th>
<th>Panel B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy ratio (AR) (%)</td>
<td>Root mean square normalized error (RMSNE)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>FIM</td>
<td>94.08</td>
<td>91.86</td>
</tr>
<tr>
<td>MLP ($\delta = 1$)</td>
<td>93.69</td>
<td>91.76</td>
</tr>
<tr>
<td>MLP ($\delta = 6$)</td>
<td>93.30</td>
<td>91.52</td>
</tr>
<tr>
<td>MLP ($\delta = 12$)</td>
<td>92.77</td>
<td>91.11</td>
</tr>
<tr>
<td>LSTM ($\delta = 1$)</td>
<td>93.87</td>
<td>92.03</td>
</tr>
<tr>
<td>LSTM ($\delta = 6$)</td>
<td>93.46</td>
<td>91.84</td>
</tr>
<tr>
<td>LSTM ($\delta = 12$)</td>
<td>92.81</td>
<td>91.27</td>
</tr>
<tr>
<td>GRU ($\delta = 1$)</td>
<td>93.54</td>
<td>91.37</td>
</tr>
<tr>
<td>GRU ($\delta = 6$)</td>
<td>93.48</td>
<td>91.91</td>
</tr>
<tr>
<td>GRU ($\delta = 12$)</td>
<td>93.03</td>
<td>91.45</td>
</tr>
<tr>
<td>Improvement (%)</td>
<td>0</td>
<td>0.19</td>
</tr>
</tbody>
</table>
Results

48-month prediction horizon
Conclusion

Multiperiod default prediction

Real world practical scenarios

Outperform the SOTA

A term structure of monotonically increasing CDP
Default occurrences
Thank you