Text Analysis

Flow indicators are constructed to forecast banking crises. Findings from Granger causality highlight leading indicators status of the Banking Crisis Lexicon Index. (see Figure above). The aggregated Sentiment Index index as a coincident indicator for developed economies is constructed from the lexicon. A combined lexicon and sentiment index act as solid forecasting indicators. Statistical models Wordscore and Wordfish are introduced to study banking crises and underscores crisis classification strength.

Thomson Reuters News archive

- Large volume featuring over 19 million news articles
- International coverage all application across multiple countries
- Newsfeeds used by broadcasters and publishers, accessible to analysts and general public
- Source is bound to the News archive
- Date spans the period 1887 to 2019
- Language is constrained to English
- Source is bound to the News archive
- Date spans the period 1887 to 2019
- Language is constrained to English
- Region includes an aggregated global configuration and separate selection of 23 individual countries, respectively 18 developed economies and 5 emerging markets

Identification and dating of banking crises

1. The closure, merger or takeover of a financial institution by the government or
2. The provision of financial assistance to a financial institution by the government (Reinhart and Rogoff, 2009)

Dimensionality reduction

- 27,406 individual terms
- Dimensionality reduction
- Pre-processing techniques
- Stop words
- Stemming
- Sparsity (0.5 percent)

Index 2: Wordscore

- The model is assumed to follow a Poisson distribution, and formally stated as:
  \[ \text{Wordscore} = \sum_{i=1}^{w} \log \left( \frac{f_i}{\text{expected}_i} \right) \]

- Wordscore is the count of word occurrences in a set of documents

Step 1: Calculate start values (\( \lambda_i \) and \( \phi \))
Step 2: Estimate document parameters (\( \lambda_i \) and \( \phi \)) and maximise log-likelihood for documents
Step 3: Estimate word parameters (\( \lambda_i \) and \( \phi \)) and maximise log-likelihood for each word
Step 4: Calculate log likelihood
Step 5: Repeat steps 2-4 until convergence

Index 3: Banking Crisis Lexicon Index

| Year | BCLI | Share of World (%)
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>2001</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>2002</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Index 4: Sentiment Index

Positive Sentiment
- certain, expect, clarity, encourage, exit, immediate, plan, attract, excel, impress, prove or good

Negative Sentiment
- uncertain, suspect, concrete, discourage, hold, poor, pass, jitter, fall, crisis, decline, guarantee, terrible, want, end, index, waste, comply, fail, woe, sharp or fear