

## TEXT ANALYSIS

Five indicators are constructed to forecast banking crises. Findings from Granger causality highlight leading indicators status of the **Banking Crisis Lexicon Index**, up to two years preceding a crisis. While the aggregated **Sentiment Index** constitutes a coincidental indicator, for developed economies it is a short-term **leading indicator**. A **combined lexicon and sentiment index** exhibit solid **forecasting performance**. Statistical models **Wordscores** and **Wordfish** are introduced to study banking crises and underscore crisis classification strength.

### Thomson-Reuters News archive

- Large volume featuring over 19 million news articles
- International coverage allows applicability across multiple countries
- Newsfeeds used by broadcasters and publishers, accessible to analysts and general public
- Source is bounded to the News archive
- Date spans the period 1987 to 2019
- Language is constraint to English
- Region include an aggregated global configuration and separate selection of 23 individual countries, respectively 18 developed economies and 5 emerging markets

Country	Crises Years	Control Years
United States	2007 - 2010	-
United Kingdom	2007 - 2014	-
Austria	2008 - 2011	-
Belgium	2008 - 2014	-
Denmark	2008 - 2014	-
France	2008 - 2014	-
Germany	2008 - 2010	-
Italy	2008 - 2014	-
Netherlands	2008 - 2014	-
Sweden	2008 - 2010	-
Canada	-	2005 - 2013
Japan	-	2005 - 2013
Greece	2008 - 2014	-
Ireland	-	2005 - 2013
Portugal	2008 - 2014	-
Spain	2008 - 2014	-
Australia	-	2005 - 2013
New Zealand	-	2005 - 2013
South Africa	-	2005 - 2013
Mexico	-	2005 - 2013
Thailand	-	2005 - 2013
Czech Republic	-	2005 - 2013
Poland	-	2005 - 2013

### Identification and dating of banking crises

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

Top 20	2004	2005	2006	2007	2008	2009	2010	2011	2012
1	bank	rate	shares	u.s	shares	bank	bank	bank	bank
2	rate	bank	market	market	high	economy	market	market	economy
3	reserve	up	bank	prices	bank	u.s	rate	rate	market
4	federal	market	european	economy	price	shares	economy	u.s	rate
5	economy	prices	up	rates	up	market	u.s	up	inflation
6	u.s	foreign	stocks	bank	market	up	china	prices	u.s
7	earnings	u.s	u.s	inflation	rise	financial	financial	inflation	policy
8	market	bonds	rate	up	european	credit	reserve	oil	central
9	issues	debt	federal	growth	u.s	rate	up	china	up
10	dollar	balance	inflation	credit	euros	federal	federal	policy	credit
11	policy	interest	reserve	financial	stocks	policy	inflation	interest	growth
12	interest	federal	earnings	dollar	federal	central	central	financial	housing
13	prices	inflation	prices	interest	inflation	debt	growth	central	financial
14	monetary	central	euros	shares	earnings	crisis	investors	down	monetary
15	trade	trade	economic	federal	rise	down	policy	investors	debt
16	balance	reserve	growth	policy	rate	reserve	credit	global	interest
17	foreign	shares	investors	high	outlook	interest	euro	federal	federal
18	house	monetary	rise	oil	reserve	global	interest	reserve	prices
19	sales	currency	price	fed	investors	inflation	debt	monetary	reserve
20	european	plans	oil	reserve	down	government	government	stocks	data



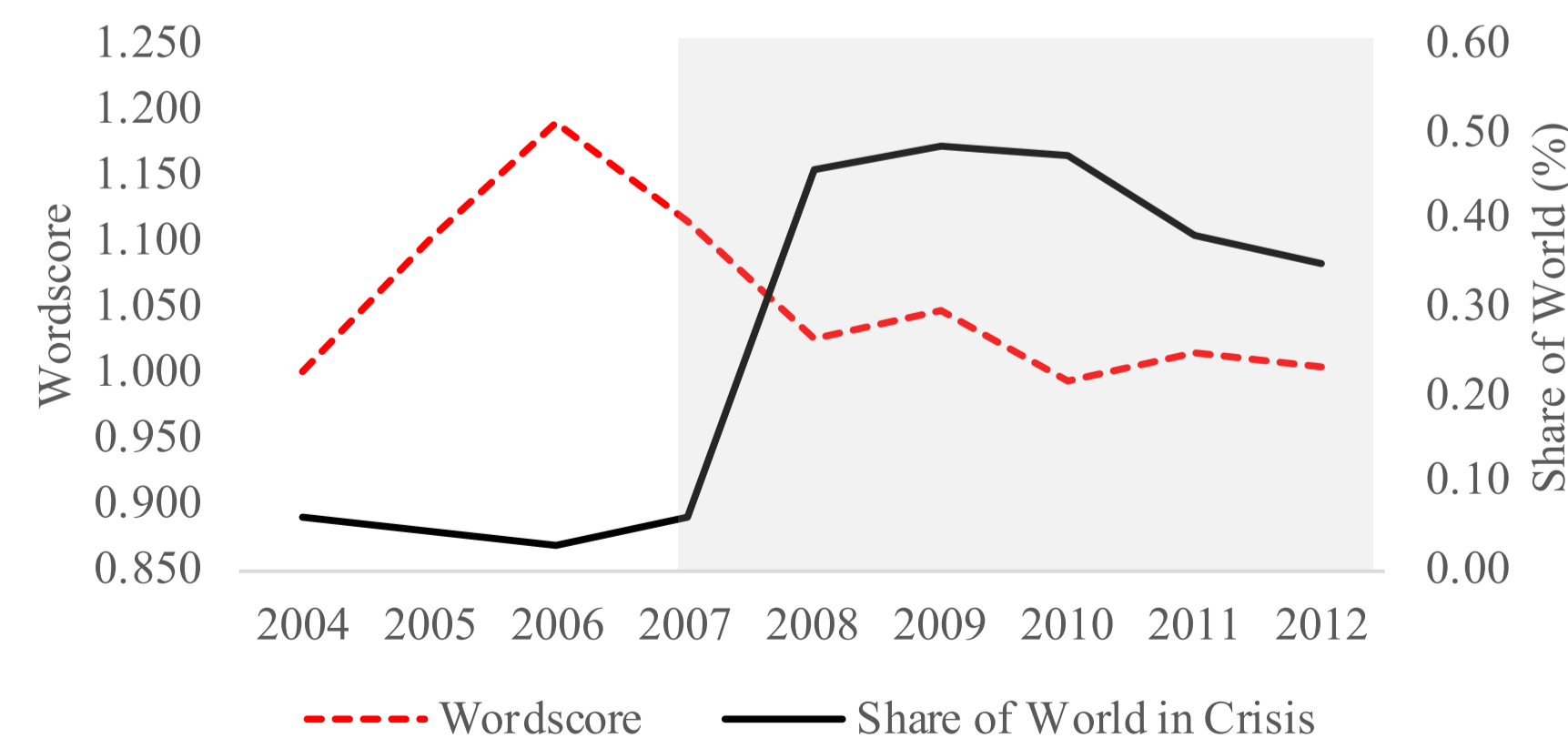
### Dimensionality reduction

- 27,408 individual terms
- Dimensionality reduction
- Pre-processing techniques
  - Stop words
  - Case folding
  - Stemming
- Sparsity (0.5 percent)

## INDEX 1: WORDSCORES

- A priori policy positions
- Compare virgin texts to reference texts
- Positions and unique words

- $A_{rd} >$  reference text  $R$ , with a priori policy position on dimension  $d$
- $F_{wr} >$  compute reflective frequency of word  $w$  as proportion of total words in reference text  $r$
- $P_{wr} = \frac{F_{wr}}{\sum_r F_{wr}} >$  use matrix of relative word frequencies to estimate conditional probabilities
- $S_{wd} = \sum_r (P_{wr} A_{rd}) >$  we can then use this matrix  $P_{wr}$  to produce a score for each word  $w$  on dimension  $d$
- $S_{vd} = \sum_w (F_{vw} S_{wd}) >$  then we must compute the relative frequency of each virgin text word, as a proportion of the total number of words in the virgin text
- $S^*_{vd} = (S_{vd} - S_{vd}) \left( \frac{SD_{rd}}{SD_{vd}} \right) + S_{vd} >$  where  $S_{vd}$  is the average score of the virgin texts,  $SD_{rd}$  and  $SD_{vd}$  are added as standard deviations of reference and virgin texts
- $V_{vd} = \sum_w F_{vw} (S_{wd} - S_{vd})^2 >$  We compute  $V_{vd}$ , the variance of each word's score around the text's total score, weighted by the frequency of the scored word in the virgin text.



## INDEX 2: WORDFISH

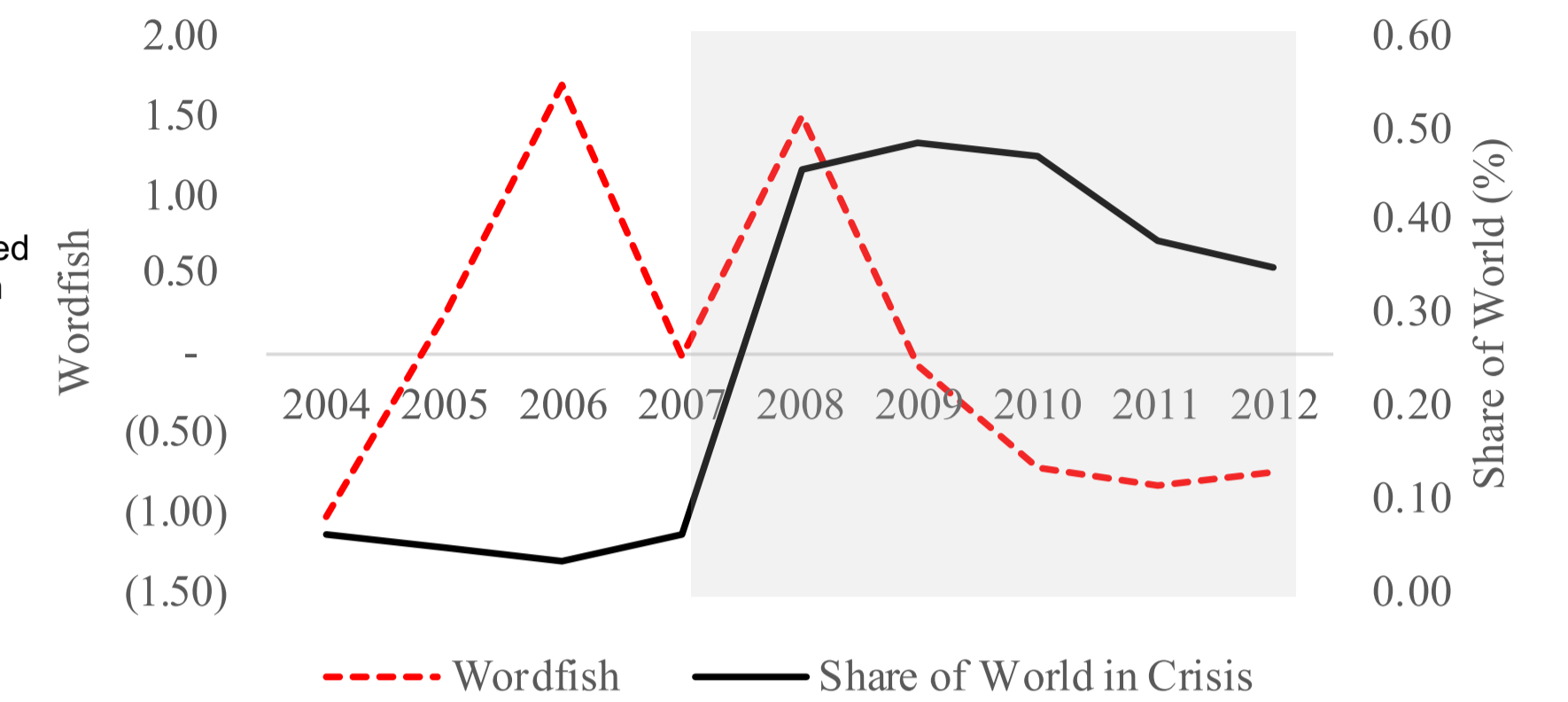
The model is assumed to follow a Poisson distribution, and formally stated as:

$$Wordcount_{ij} \sim Poisson(\lambda_{ij})$$

$$\lambda_{ij} = \exp(\alpha_i + \psi_j + \beta_j \omega_i)$$

Wordcount is the count of word  $j$  in text  $i$ , with  $\alpha$  a set of document fixed effects,  $\psi$  as word fixed effects,  $\beta$  represents an estimated word specific weight to highlight the importance of word  $j$  in distinguishing between positions, and  $\omega$  is an estimate of document  $i$ 's position

- Step 1: Calculate starting values  $(\alpha_i + \psi_j + \beta_j \omega_i)$
- Step 2: Estimate document parameters  $(\alpha_i + \omega_i)$  & maximise log-likelihood for documents
- Step 3: Estimate word parameters  $(\psi_j + \beta_j)$  & 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

### Banking Sector

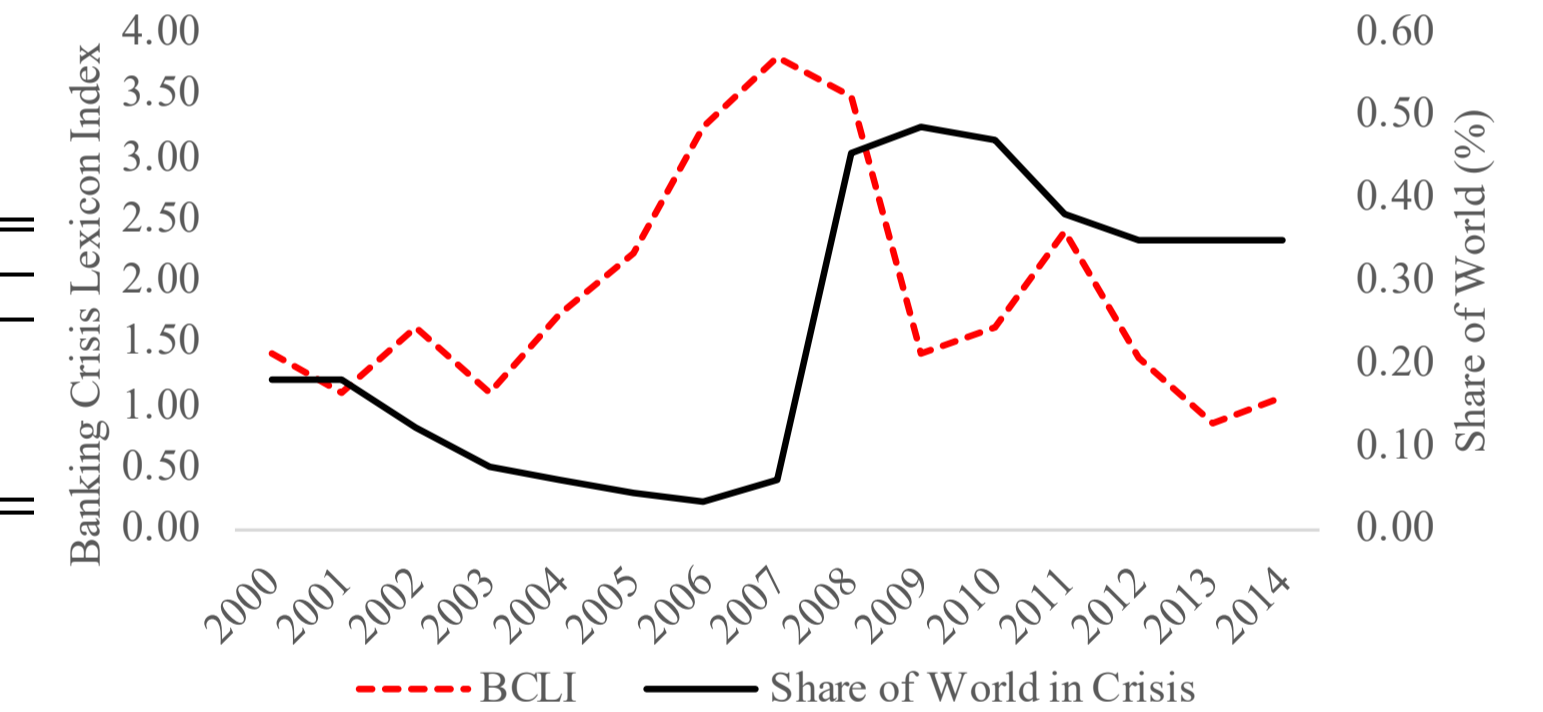
bank or banking  
banking  
deposit or credit OR debt  
interest rate  
inflation or cpi  
reserve or gold  
liquid or contract or eas or tight or monetary or boom or bust or crisis  
fraud or earning or hous

$$BCLI = \frac{((B_1+B_2+\dots+B_M)_t \cdot [R_1+R_2+\dots+R_M]_t \cdot [E_1+E_2+\dots+E_M]_t)}{(\sum_{i=1}^M A_i)}$$

Lag	World		Developed Countries		Emerging Markets	
	BCLI	Crisis	BCLI	Crisis	BCLI	Crisis
1	7.686***	4.685**	8.597***	8.269***	2.267	0.471
2	4.873**	1.851	4.115**	4.983**	2.610	0.089
3	3.987*	3.009	2.764	2.326	2.000	0.277
4	46.074**	3.892	4.112	9.547*	0.866	1.732

### Real Sector

consum or invest or produc  
Export or import or trade or terms of trade



## INDEX 4: SENTIMENT INDEX

$$Sentiment Index = \frac{(Sentiment_n - Sentiment_p)}{(Sentiment_n + Sentiment_p)}$$

where  $Sentiment_n$  = total text units in category: negative sentiment  
and  $Sentiment_p$  = total text units in category: positive sentiment.

### Positive Sentiment

bank and  
certain, expect, clarity, encourage, excit, incredible, pleas, attract, excel, impress, postiv or good

### Negative Sentiment

bank and  
cnertain, unexpect, concern, discouragement, bad, poor, panic, jitter, fail, crisis, distrust, jeopardy, terribl, worry, erod, reduc, warn, complicat, fear, woes, slump or low

