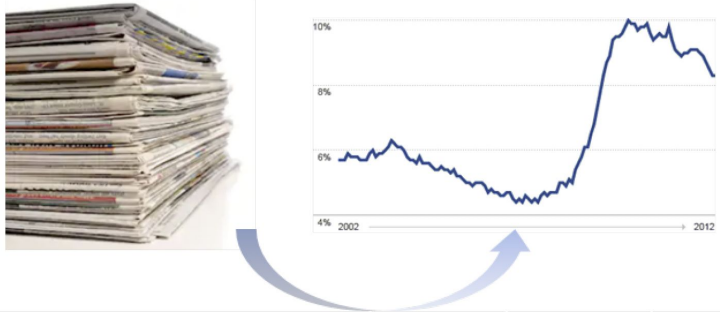
Measure Country-Level Socio-Economic Indicators with Streaming News: An Empirical Study

Bonan Min, Xiaoxi Zhao

Motivation

- Social-economic indicator
  - Measuring economic conditions
  - Input for policy maker (e.g. president)
- Very difficult to obtain
  - Expensive (cencus is usually conducted every 10 year)
  - Inaccurate (e.g.unemployment rate)
  - Laggy
- Hypothesis
  - Correlation between events in streaming news and economic events.
  - Streaming news can be used to measure these indicators
  -

Framework



Event triggers	Location	Time	Related*
the U.S. **recession** began in December 2007	U.S. (USA)	December 2007 (2007-12)	Y
SEPT. 1, 2011 The **bankruptcies** of three American solar power companies in the last month	American (USA)	Last month (2011-08)	Y
The man accused of fatally **shooting** his estranged wife inside a New Jersey church last Tuesday	New Jersey (USA)	Last Tuesday (2009-07)	N
…	…	…	

- Data: English Gigaword corpus (1997-2010)

Extract events

1. Label predicates using SRL (He et al 2017)
2. Extract predicates as triggers, entity, time, location mentions
3. Refine location into country-level GPE
4. Refine time entity (absolute and relative time entities)

<clashed, Syria, 2018-11-23>

Find relevant events

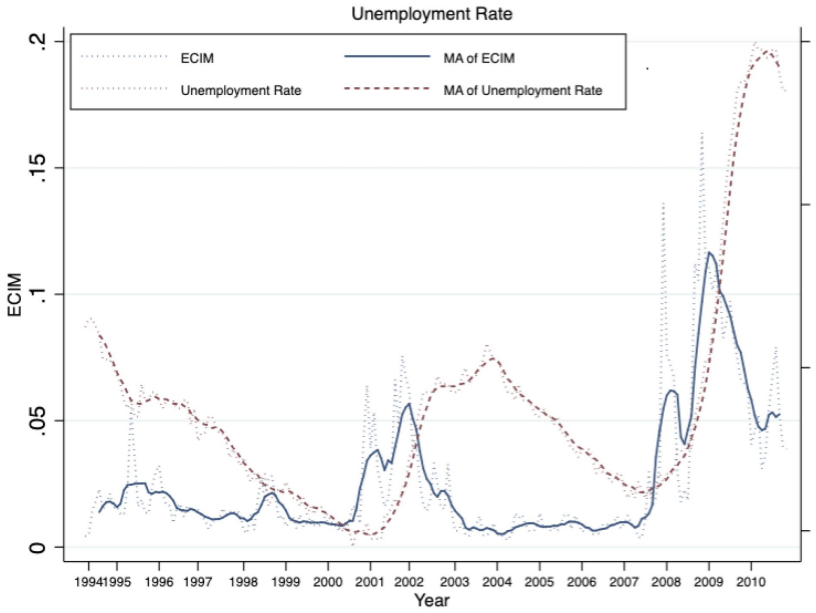
- Process 500 documents (as above)
- Annotate trigger words for target indicators
- Using wordnet to extend the word list
- Refine word list by another annotator

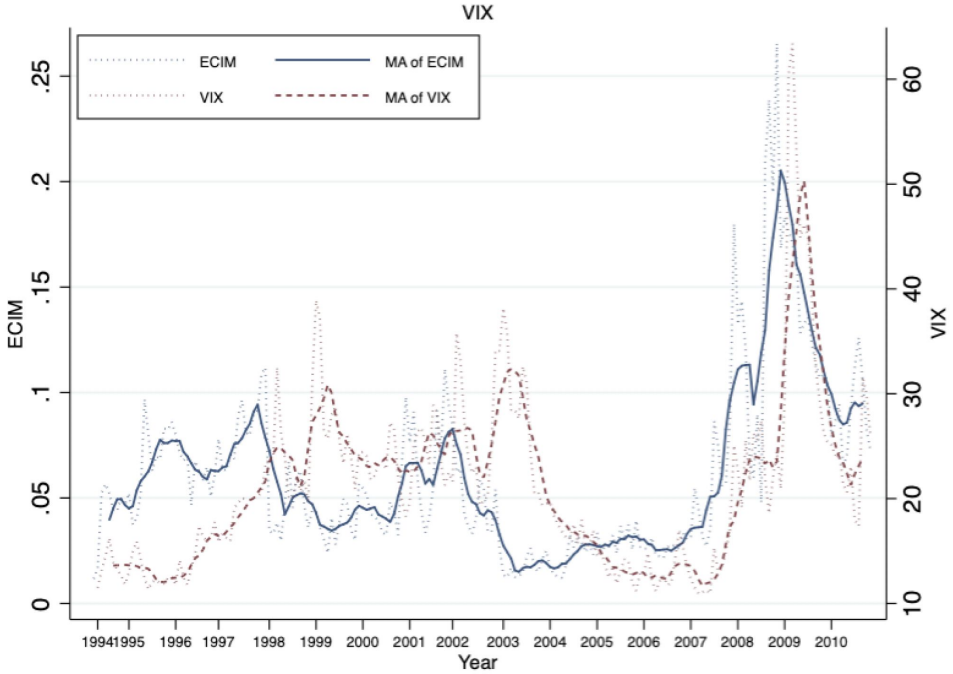
Measure indicator

1. Aggregate and count in 1-month bin
2. Normalize
3. Average smoothing (7 months)

$$ECIM_{i,t} = \frac{1}{T} \sum_{t' \in [t-\frac{T}{2}, t+\frac{T}{2}]} \frac{\sum_{e \in \mathbb{E}_i} N_{e,t'}}{M_{t'}}$$

Results


Unemployment Rate
ECIM
MA of ECIM
Unemployment Rate
MA of Unemployment Rate
ECIM
0
.05
.1
.15
.2
1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
Year


VIX
ECIM
MA of ECIM
VIX
MA of VIX
ECIM
0
.05
.1
.15
.2
.25
VIX
10
20
30
40
50
60
1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
Year

Correlation test

Pearson correlation test: To test the confidence of the correlation between two variables

Indicator	**Pearson**	**P-value**
Unemployment rate $(1/\mathbb{M}_{comp})$	0.4286 (0.4877)	0.0000 (0.0000)
EPU News	0.5136	0.0000
VIX	0.4115	0.0000

Table 2: Correlation coefficients between ECIMs and indicators. For unemployment rate, we also show its correlation with $1/\mathbb{M}_{comp}$ (in parentheses).