Extracting Socioeconomic Patterns from the News: Modelling Text and Outlet Importance Jointly

Vasileios Lampos[β]
v.lampos@ucl.ac.uk

Daniel Preoţiuc-Pietro[δ]
danielpr@gmail.com

Sina Samangooei[σ]
sinjax@gmail.com

Douwe Gelling[γ]
d.gelling@sheffield.ac.uk

Trevor Cohn[τ]
t.cohn@unimelb.edu.au

[β] UCL
[δ,γ] The University Of Sheffield.
[σ] University of Southampton
Trend Miner
[τ] The University of Melbourne

2006 -> 2013

Data

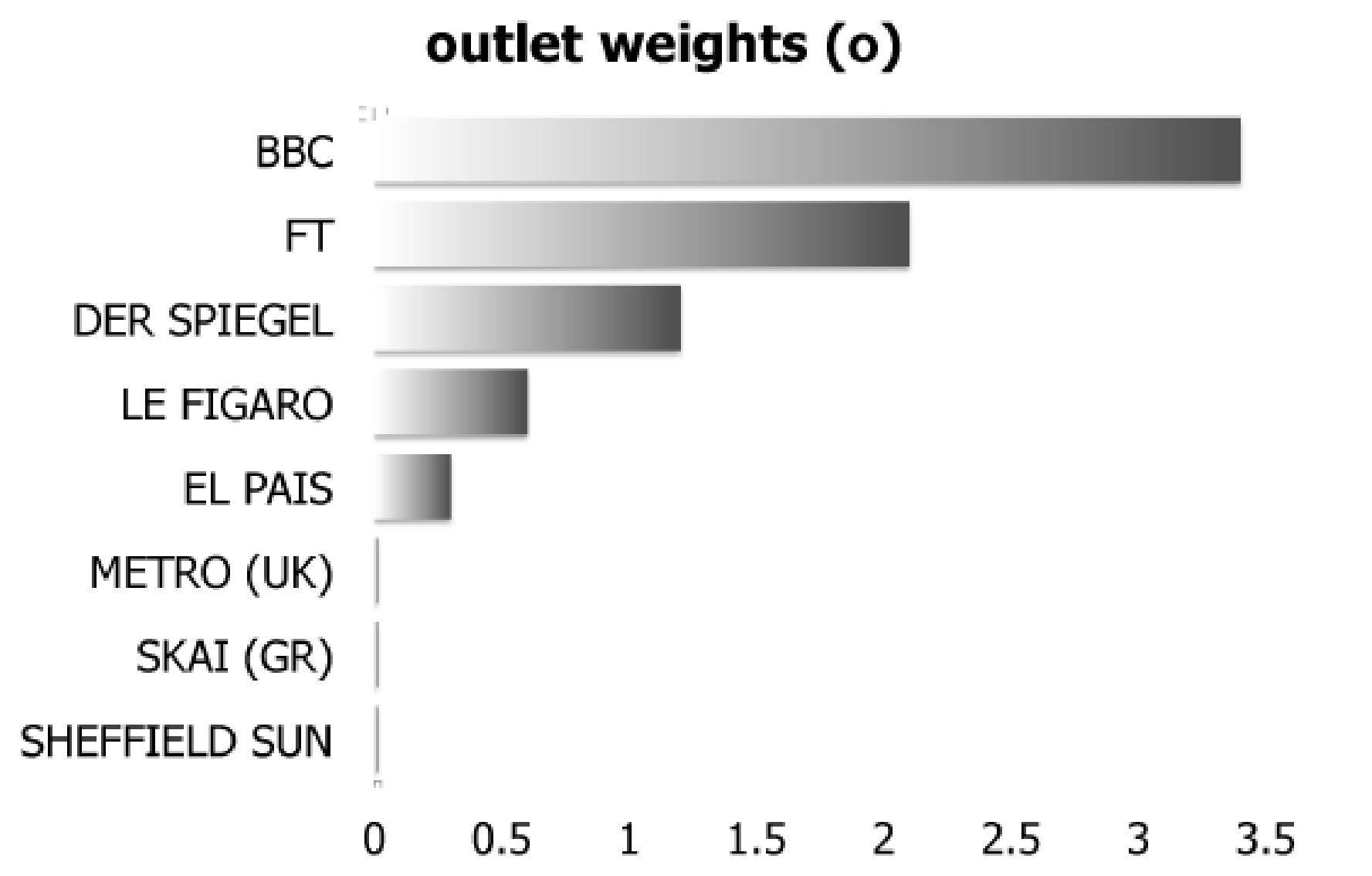
- news summaries written in English
- extracted from the **Open Europe Think Tank**
- daily aggregat1ion of news about the EU
- focused on current affairs, politics, economy
- each summary is a few paragraphs long
- each summary is attached to ≥1 news outlets
- **1913 days**, 94 months
- avg. 14 news/summary
- features: **8413 unigrams** + **19045 bigrams**
- **435 news outlets**

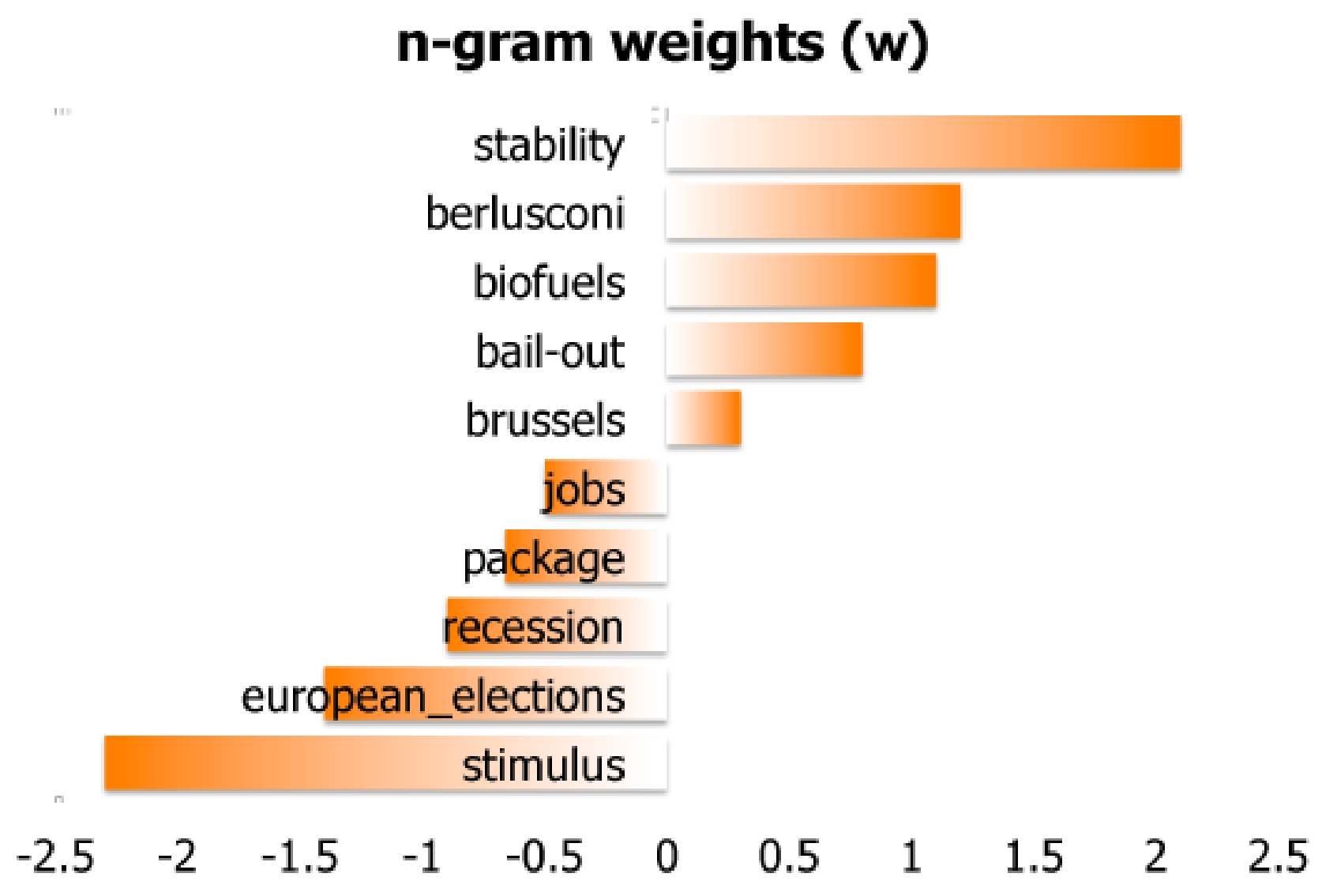
$$f \; : \; X \; \rightarrow \; Y$$

{(, OUTLET)} ESI*

prediction — outlet x word daily frequencies — bias term

$$y = o\,X\,w + \beta$$

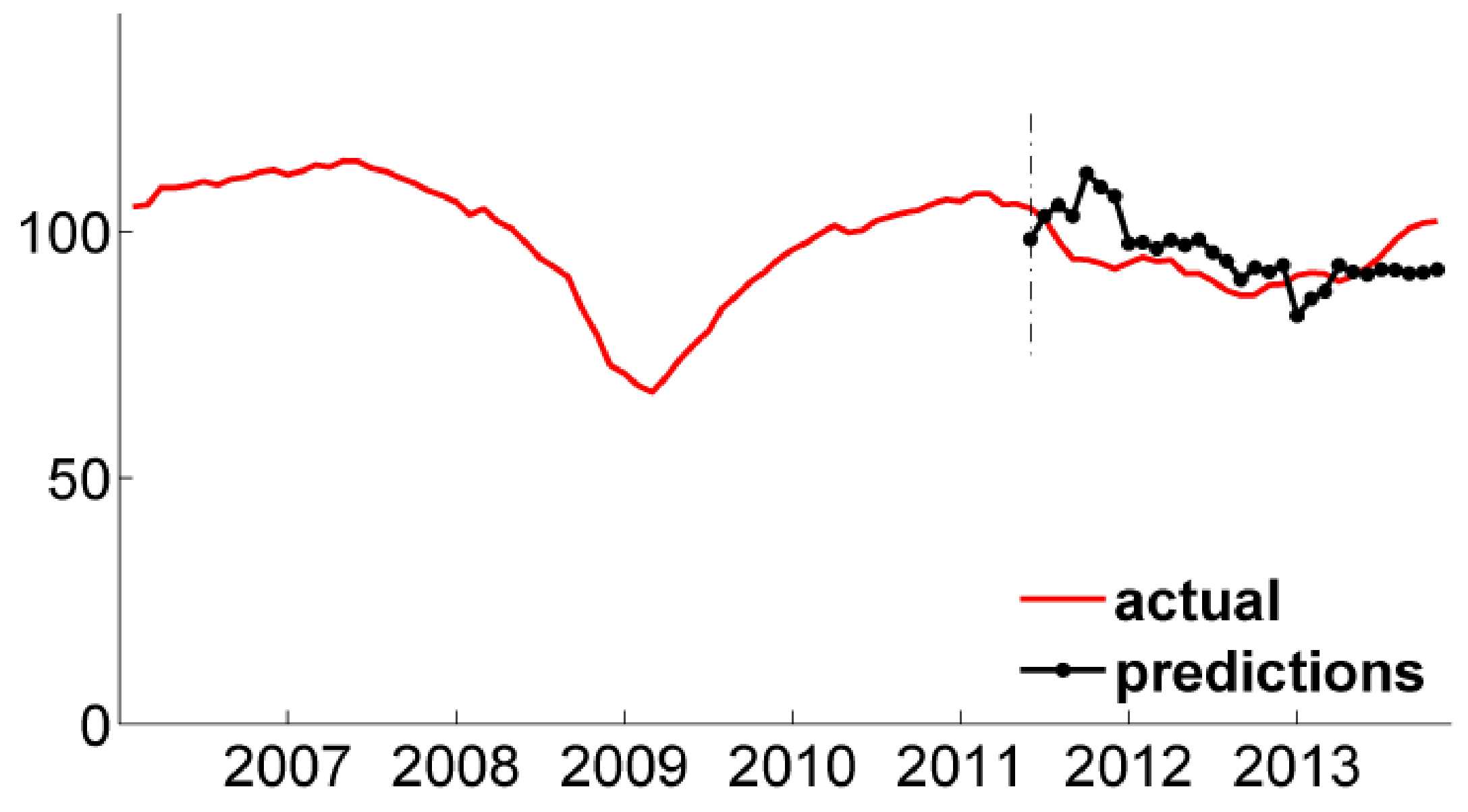




***ESI = Economic sentiment indicator**

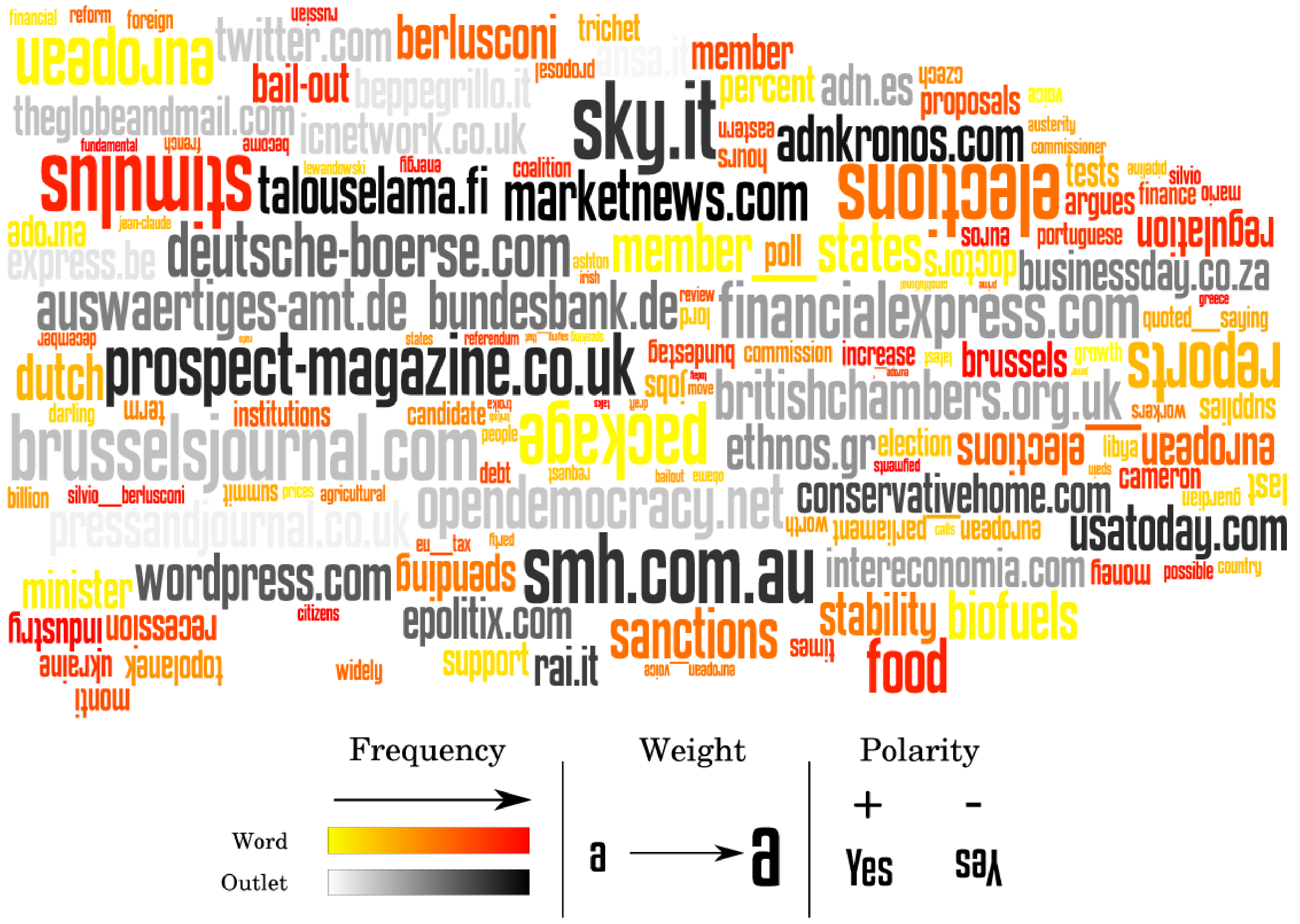
- composite indicator often seen as an early predictor for future economic developments
- consists of five confidence indicators with different weights: industrial (40%), services (30%), consumer (20%), construction (5%) and retail trade (5%)
- any socioeconomic indicator can be used in this framework

Forecast



Linear regression: 9.253 (9.89%)
Bilinear reression: 8.209 (8.77%)

RMSE (error rate %) in a forecasting setup on 30 data points


financial
reform
foreign
russian
european
twitter.com
berlusconi
trichet
member
bail-out
beppegrillo.it
ansa.it
proposal
percent
adn.es
czech
proposals
theglobeandmail.com
icnetwork.co.uk
sky.it
eastern
adnkronos.com
austerity
commissioner
fundamental
french
became
stimulus
talouselama.fi
lewandowski
energy
coalition
marketnews.com
elections
tests
silvio
finance
argues
euro
regulation
jean-claude
express.be
deutsche-boerse.com
ashton
irish
member
poll
states
doctors
portuguese
businessday.co.za
auswaertiges-amt.de
bundesbank.de
review
financialexpress.com
greece
quoted__saying
december
states
referendum
dutch
prospect-magazine.co.uk
budgets
commission
increase
brussels
growth
reports
jobs
move
britishchambers.org.uk
darling
term
institutions
candidate
troika
workers
supplies
brusselsjournal.com
people
package
ethnos.gr
election
elections
libya
european
debt
request
bailout
obama
cameron
billion
silvio__berlusconi
summit
prices
agricultural
opendemocracy.net
conservativehome.com
guardian
last
pressandjournal.co.uk
worth
parliament
calls
european
usatoday.com
eu__tax
fixed
minister
wordpress.com
spending
smh.com.au
intereconomia.com
money
possible
country
citizens
epolitix.com
industry
recession
sanctions
stability
biofuels
times
ukraine
topolanek
widely
support
rai.it
european__voice
food
monti
Frequency
Weight
Polarity
Word
Outlet
a
a
Yes
Yes