

Workshop on using big data for forecasting and statistics

Monday, 7 and Tuesday, 8 April 2014
European Central Bank, Eurotower
Frankfurt am Main





EUROPEAN CENTRAL BANK

EUROSYSTEM

[Please select]

[Please select]

Dr. Piet J.H. Daas
Methodologist & Research
coordinator Big Data
Statistics Netherlands

Dr. Marco J.H. Puts
Senior Statistical
Researcher
Statistics Netherlands

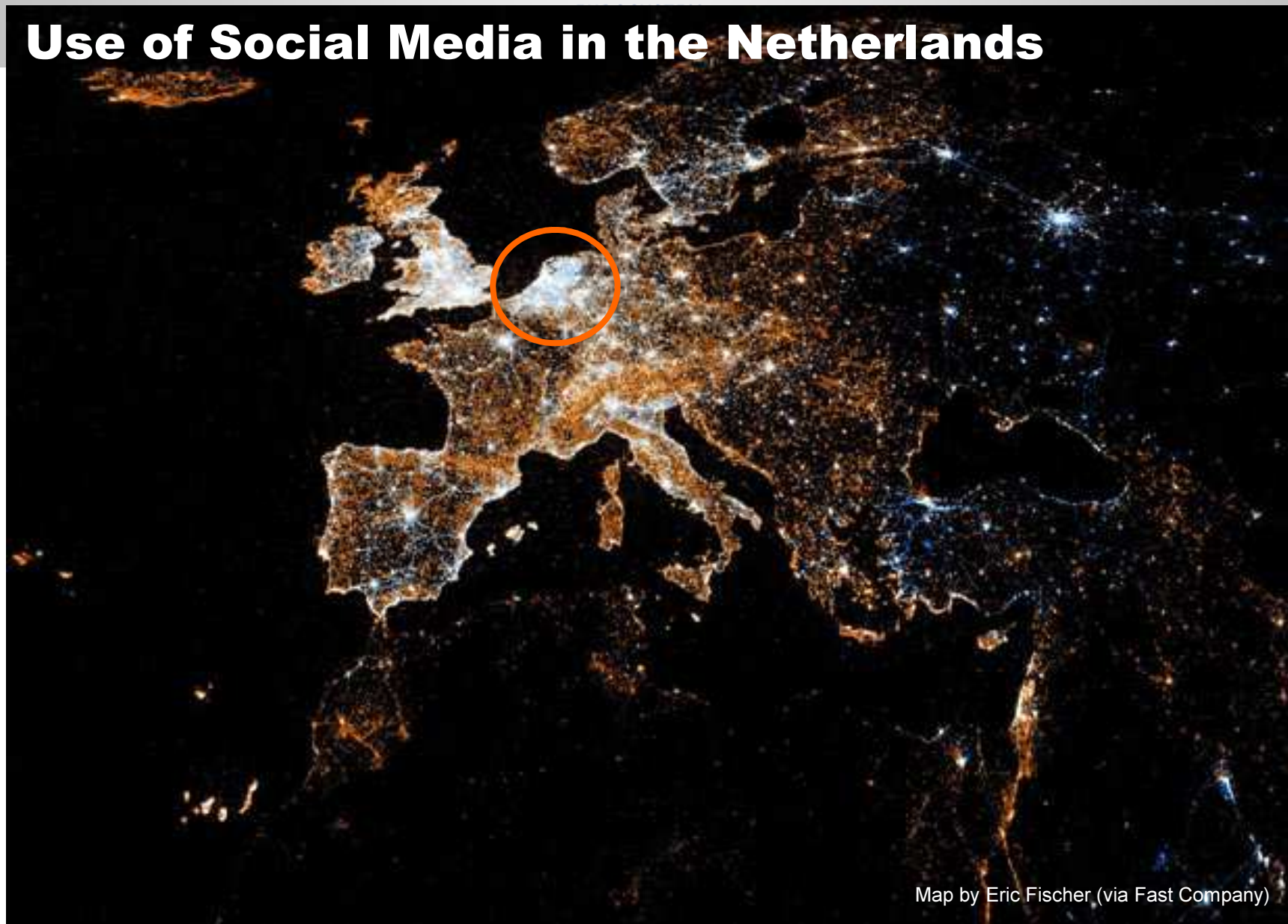
Social Media Sentiment and Consumer Confidence

From a statistical point of view



EUROPEAN CENTRAL BANK

Use of Social Media in the Netherlands



Map by Eric Fischer (via Fast Company)

Data source and sentiment determination

- About the data
 - Dutch firm that continuously collects *ALL public* social media messages written in Dutch



- Dataset of more than 3.5 billion messages!
 - Covering June 2010 till the present
 - Between 3-4 million new messages are added per day

- About sentiment determination

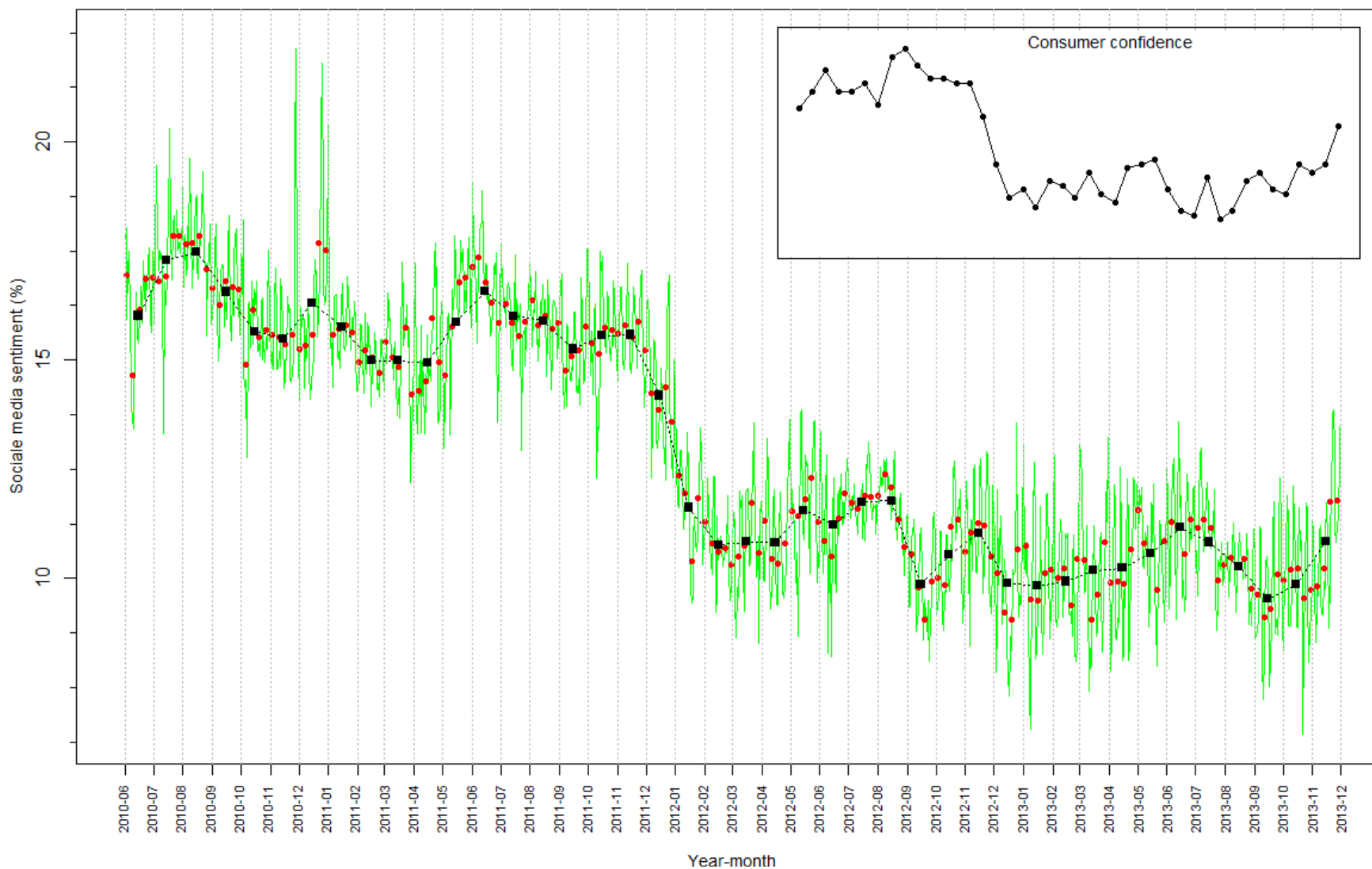
- ‘Bag of words’ approach
 - List of Dutch words with their associated sentiment
 - Added social media specific words (‘FAIL’, ‘LOL’, ‘OMG’ etc.)
- Use overall score to determine sentiment
 - Is either positive, negative or **neutral**
- Average sentiment per period (day / week / month)
 - $(\#positive - \#negative) / \#total * 100\%$



EUROPEAN CENTRAL BANK

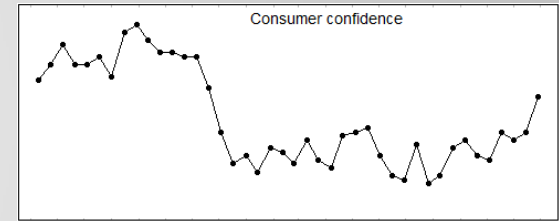
EUROSYSTEM

Daily, weekly, monthly sentiment



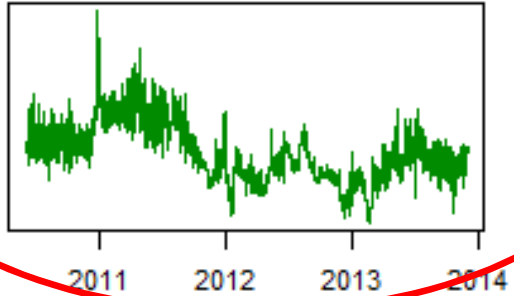


EUROPEAN CENTRAL BANK
EUROSYSTEM

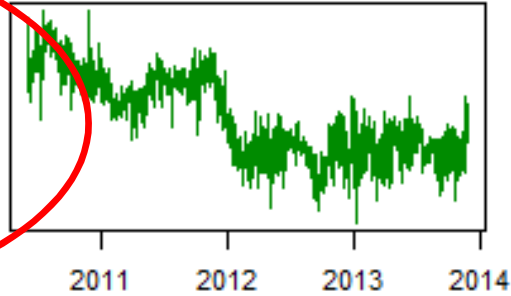


Per platform, daily

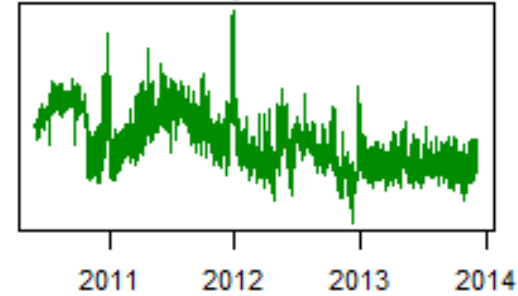
Facebook (~10%)



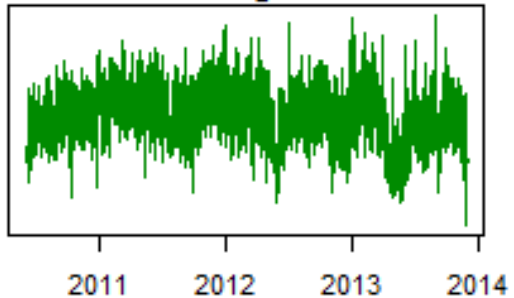
Twitter (~80%)



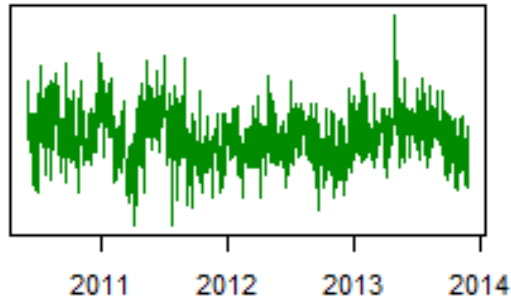
Hyves



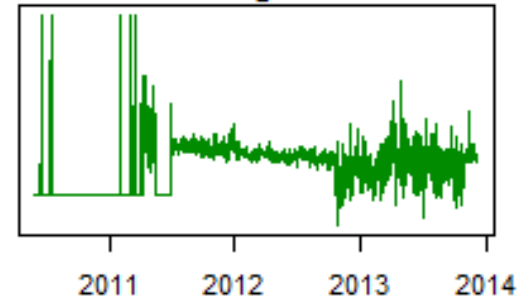
Blogs



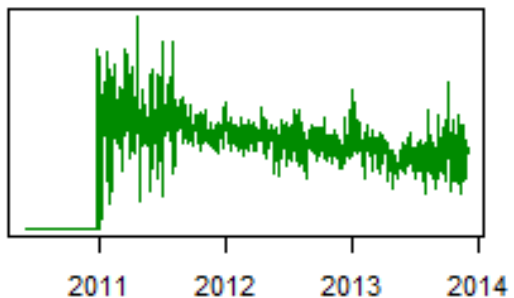
News sites



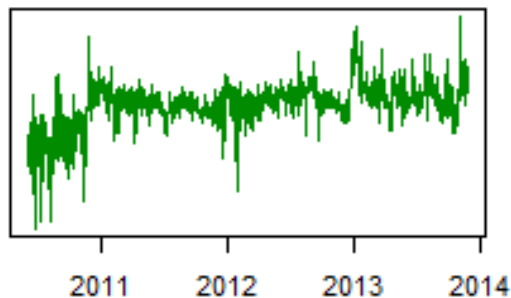
Google+



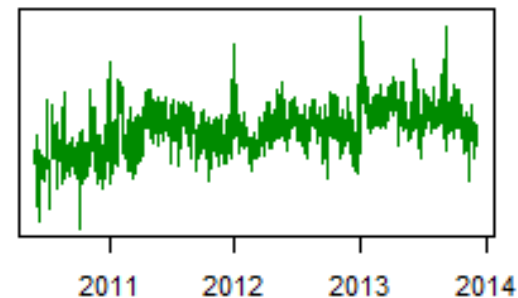
LinkedIn



Youtube



Forums



Platform specific results

Table 1. Social media messages properties for various platforms and their correlation with consumer confidence

Social media platform	Number of social media messages ¹	Number of messages as percentage of total (%)	Correlation coefficient of monthly sentiment index and consumer confidence (r) ²
All platforms combined	3,153,002,327	100	0.75
Facebook	334,854,088	10.6	0.81* ←
Twitter	2,526,481,479	80.1	0.68
Hyves	45,182,025	1.4	0.50
News sites	56,027,686	1.8	0.37
Blogs	48,600,987	1.5	0.25
Google+	644,039	0.02	-0.04
Linkedin	565,811	0.02	-0.23
Youtube	5,661,274	0.2	-0.37
Forums	134,98,938	4.3	-0.45

¹period covered June 2010 until November 2013

²confirmed by visual inspecting scatterplots and additional checks

*cointegrated



EUROPEAN CENTRAL BANK
EUROSYSTEM

Schematic overview

Previous month

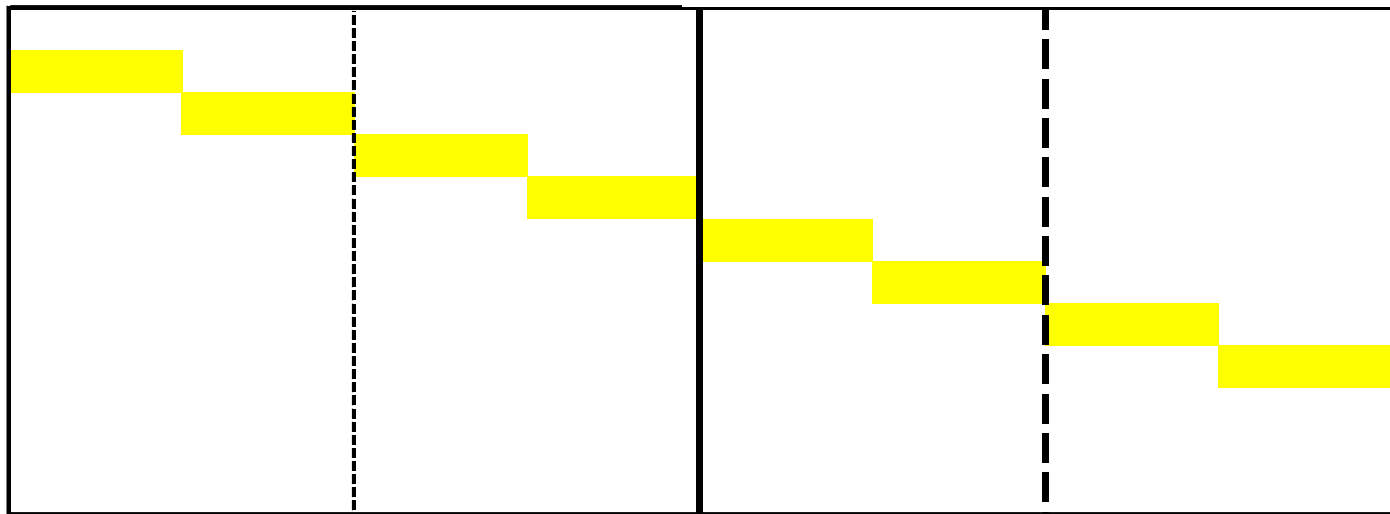
Current month

Day 1-7

Day 8-14

Day 15-21

Day 22-28



Consumer Confidence

Publication date (~20th)

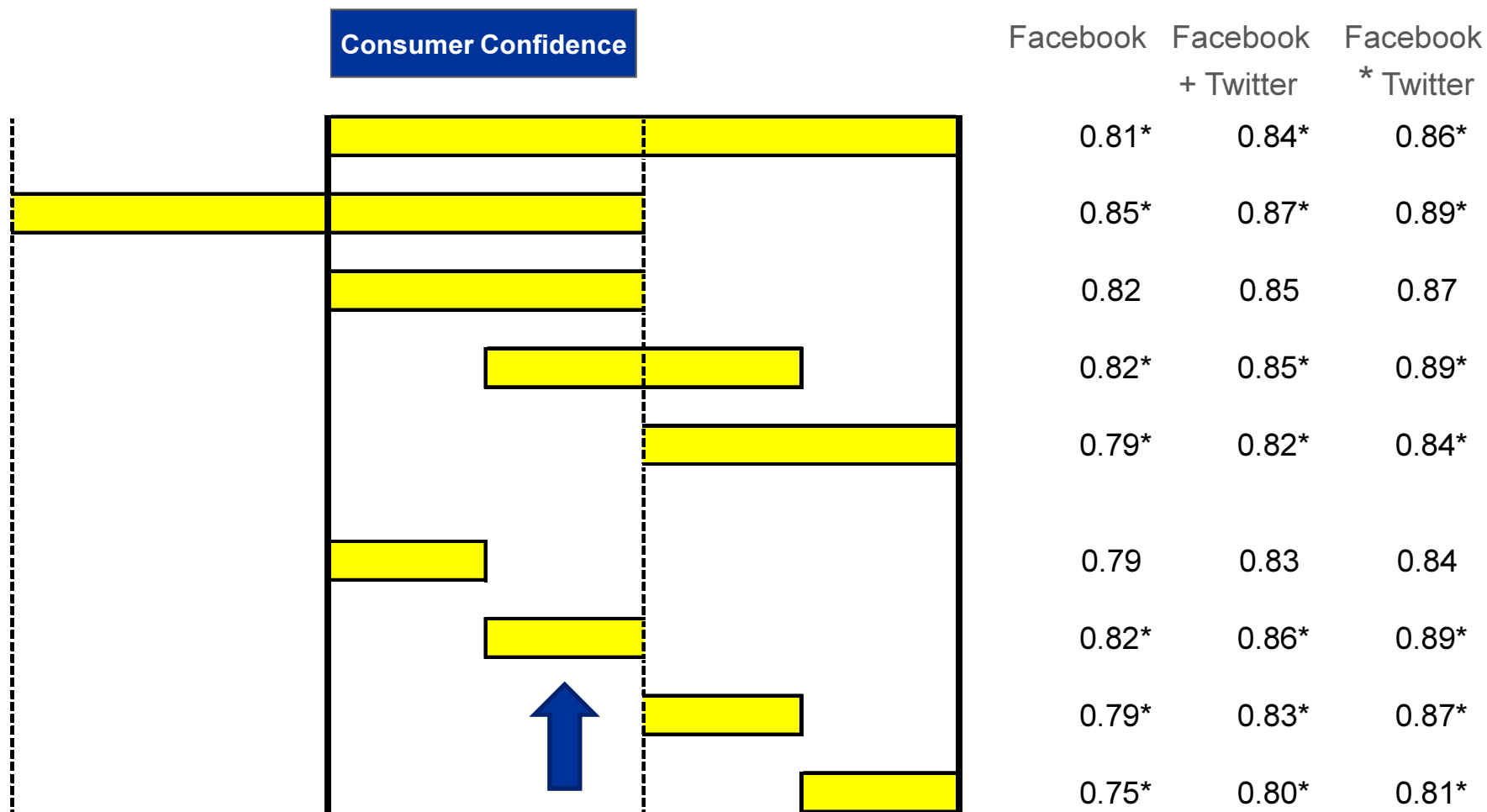
Social media sentiment



EUROPEAN CENTRAL BANK

EUROSYSTEM

Results of comparing various periods



*cointegration

LOOCV results

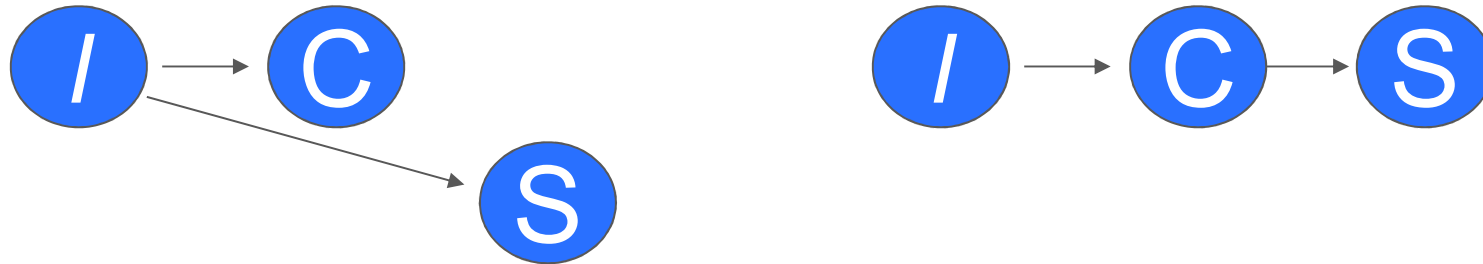


Overall findings

- Correlation and cointegration
 - 1st 'week' of Consumer confidence usually has 70% response
 - Best correlation and cointegration with 2nd 'week' of the month
 - Highest correlation 0.93* (all Facebook * specific word filtered Twitter)
- Granger causality
 - Changes in Consumer confidence *precede* changes in Social media sentiment
 - For all combinations shown!
 - Only tried linear models so far (?nonlinear relation in Twitter?)
- Prediction
 - Slightly better than random chance
 - max forecast skill score 0.12
 - For 4th 'week' of month

Overall findings (2)

- How are Confidence and Sentiment related?



- ‘Mood of the nation’ and the integral emotion in the Appraisal-Tendency Framework
- Basis for a rapid sentiment based indicator
 - Could be produced every month on or shortly after the 15th day
 - Could even be produced on a weekly basis (Volatile!)
- But, what am I comparing here!
 - *Units* differ for Consumer confidence and Sentiment
 - Confidence: Response of a representative sample of households
 - Sentiment: Messages produced during a specific period
 - ?Representatives? future studies will focus on this phenomenon!



EUROPEAN CENTRAL BANK

EUROSYSTEM

Thank you for your attention !

The
future
of
statistics!

@pietdaas

