



EUROPEAN CENTRAL BANK

EUROSYSTEM

# 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]

**Ellen Tobback**

Department of Engineering Management,  
University of Antwerp  
Speaker

**Walter Daelemans**

Department of Linguistics, University of  
Antwerp

**Hans Naudts**

DG ECFIN, European Commission

**Eric Junqué de Fortuny**

Department of Engineering Management,  
University of Antwerp

**David Martens**

Department of Engineering Management,  
University of Antwerp

# Belgian Economic Policy Uncertainty Index: Improvement through text mining

# Problem setting

## News article about economic policy uncertainty?

1. Naive method: Yes, contains the words *uncertainty, economy* and *fiscal*
2. Modality annotation
3. Text classification

### 'Grexit' Fears Unsettle Markets as Stocks Drop amid Eurozone Uncertainty

There is still no end to the political crisis in Greece, with no government having been formed. It is not even clear if Greece will pay out on a bond maturing on 15 May.

**Uncertainty** has unsettled the markets, with all the major stock markets falling and the euro weakening.

Nothing conclusive has emerged from the day and the markets will enter tomorrow once again unsure of what exactly is going to happen to the eurozone and if and when Greece will leave the single currency.

There has been much focus on the current crises in Spain and Greece, but Italy shouldn't be forgotten. It has flirted with the 7 percent "unsustainable" yield benchmark on its 10-year bonds in the past and as the eurozone's third largest **economy** a worse than expected contraction could open up a new front in the single currency area's battle to survive.

Schaeuble says euro stability must be maintained by both monetary and **fiscal** means.



# Shortcomings of the method

Two main shortcomings of this method:

1. Every article that meets the search criteria is added to the index

**Historically large amount of poppy fields in Afghanistan**

2. Articles that address EPU without mentioning the necessary keywords are excluded from the index

**Spanish domino pieces totter due to Greek crisis**



# Problem setting

## News article about economic policy uncertainty?

1. Naive method : Yes,  
contains the words  
*uncertainty, economy*  
and *fiscal*

2. Modality annotation  
Yes, modality score \_\_\_\_\_  
is larger than the cutoff-value

3. Text classification  
Yes, linear model has a  
positive output

### 'Grexit' Fears Unsettle Markets as Stocks Drop amid Eurozone Uncertainty

There is still **no** end to the political crisis in Greece, with **no** government having been formed. It is **not even clear if** Greece will pay out on a bond maturing on 15 May.

**Uncertainty** has unsettled the markets, with all the major stock markets falling and the euro weakening.

**Nothing conclusive** has emerged from the day and the markets will enter tomorrow once again **unsure** of what exactly is going to happen to the eurozone and **if** and when Greece will leave the single currency.

There has been much focus on the current crises in Spain and Greece, but Italy shouldn't be forgotten. It has flirted with the 7 percent "unsustainable" yield benchmark on its 10-year bonds in the past and as the eurozone's third largest economy a worse than **expected** contraction **could** open up a new front in the single currency area's battle to survive.

Schaeuble says euro stability must be maintained by both monetary and fiscal means.



# Document classification

- Unstructured textual data to structured data
  - Bag of words approach



Article	Greece	eurozone	crisis	increased	output	Rupo	...	EPU?
A1	1	1	1	0	0	0		Yes (1)
A2	0	0	0	1	1	0		No (-1)
A3	0	0	1	0	0	1		Yes (1)



**A1:** The eurozone crisis is escalating, especially around Greece. A euro area break-up would be chaotic..

**Ax**      1

**A2:** US manufacturing output increased more than expected. The economic activity in the world's leading economy is picking up...

**Ax**      0

**A3:** The political impasse in Belgium could take the country into a crisis. Formateur Jo Di Rupo offered to resign...

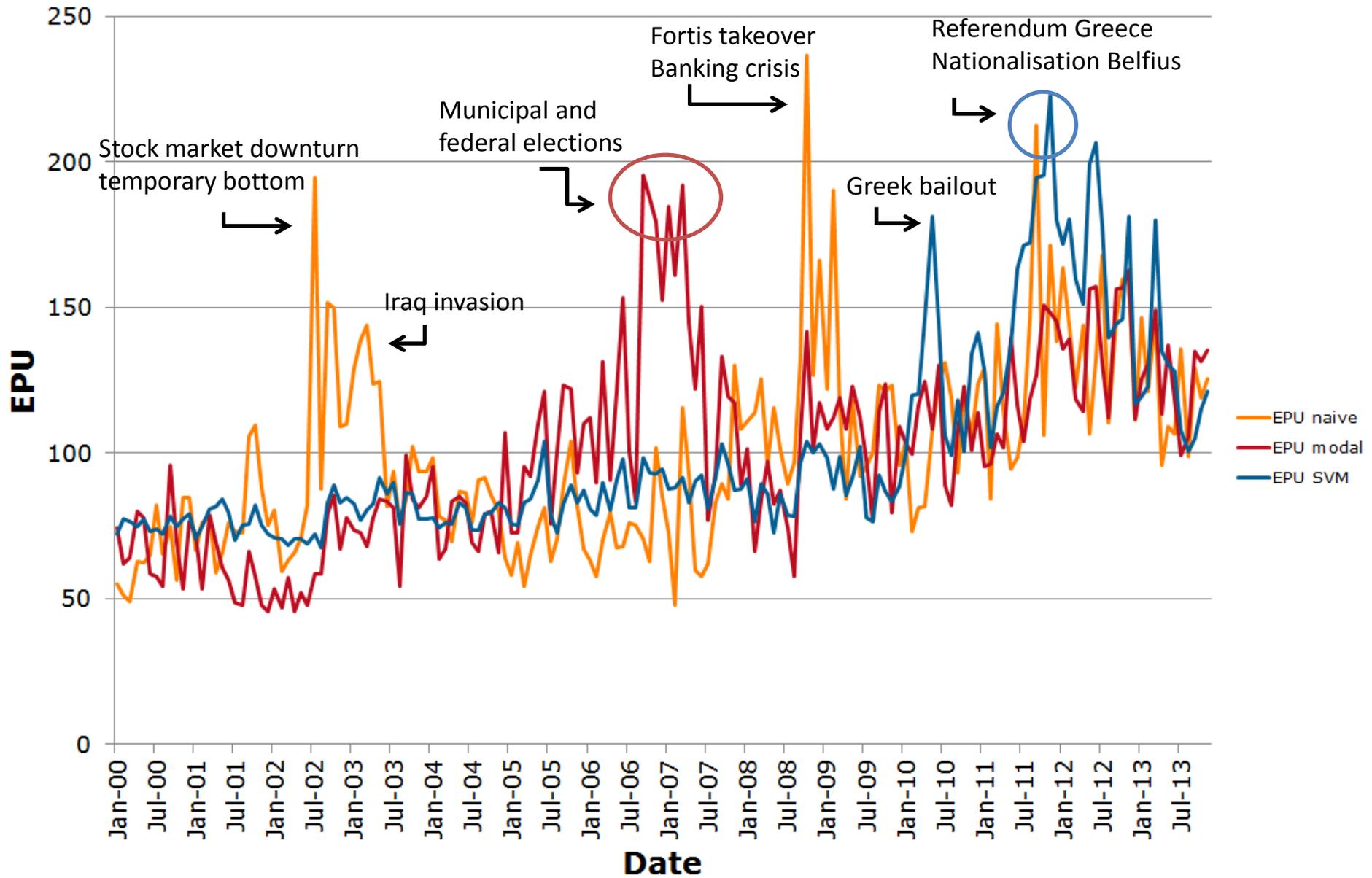
**Ax**      1

**Accuracy 96%**

**YES**

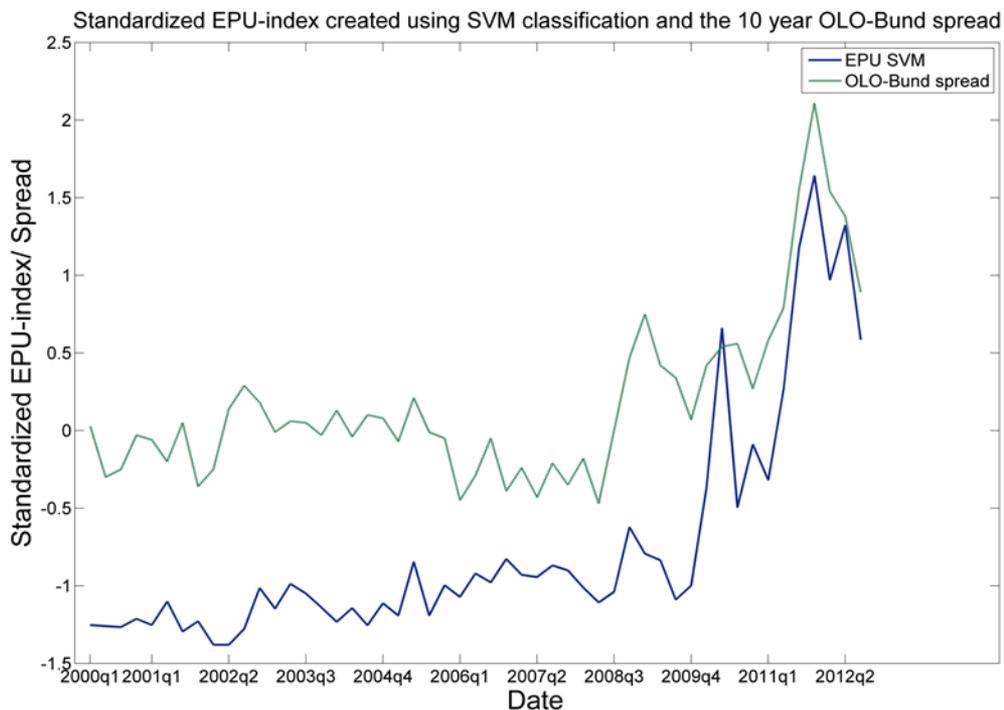
# Discriminating words





# Correlations

	EPU naive	EPU modality	EPU SVM	OLO-Bund spread
Q2 (consumer)	0.214	0.051	0.203	0.264
Q4 (consumer)	0.637	0.267	0.6033	0.738
Q5 (producer)	0.539	-0.033	0.45	0.609
OLO-Bund Spread	0.662	0.360	<b>0.872</b>	



# Conclusion

- Each method selects different articles  
→ three different EPU indicators
- Advantages of using text classification
  - Isolating country-specific EPU allows for scope refinement
  - Irrelevant articles unlikely to influence the EPU-index
  - Discriminating words defined by the model
- Towards automated tool for real-time nowcasting of macroeconomic variables based on online news



EUROPEAN CENTRAL BANK  
EUROSYSTEM

# Questions? Comments?



[Contact information](#)

[ellen.tobback@uantwerpen.be](mailto:ellen.tobback@uantwerpen.be)