

BIG DATA: USING GOOGLE SEARCHES TO PREDICT THE UNEMPLOYMENT RATE IN THE EU

AIECE MEETING BRUSSELS 6 NOV 2015

JOONAS TUHKURI, ETLA, THE RESEARCH INSTITUTE OF THE FINNISH ECONOMY
AND THE UNIVERSITY OF HELSINKI

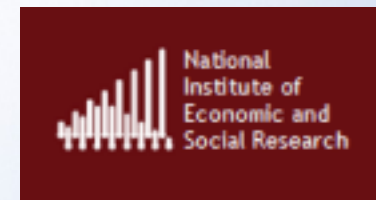
ETLA



UNIVERSITY OF HELSINKI

PARTNERS

25 Research Institutions from Europe:



100 B

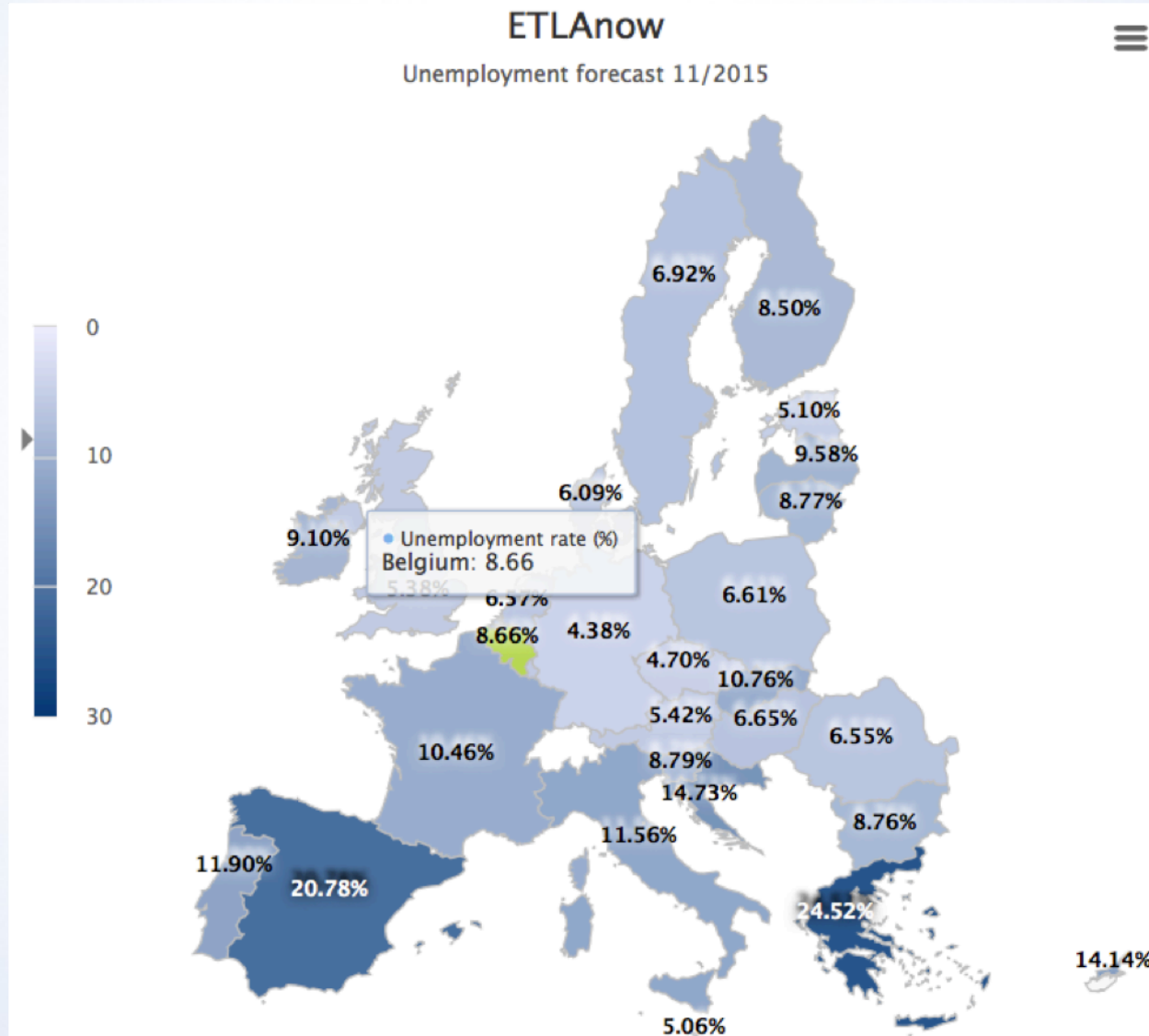
LITERATURE

- Unemployment rate (Varian & Choi 2012, Askatas & Zimmerman 2009, Tuhkuri 2014)
- Housing market (Brynjolfsson & Wu 2013)
- Sales (Goel et al 2010, PNAS)
- Macro indicators (Koop & Onorante 2013)
- Stock market (Preis et al 2013)
- Consumption (Vosen & Schmidt 2012)
- Influenza (Ginsberg et al. 2009, Nature)

ETLA NOW

<https://www.etla.fi/en/etlanow-eu28/>
Username and password: etlanow2015

ETLAnow Maps



ETLAnow Forecasts

Belgium (BE)

Unemployment rate (%)	8/2015	9/2015	10/2015	11/2015	12/2015	1/2016	2/2016
Official	9.00	8.80
ETLAnow	8.87	8.92	8.72	8.66	8.63	8.62	8.53
Change (pp)			+0.12	+0.16	+0.03	-0.18	-0.27
							Date: 4.11.2015

. = official data not available

Last update 4.11.2015.

Next update 5.11.2015.

Next official release on the unemployment rate 30.10.2015.

Export [ETLAnow forecasts for this table](#) or [all ETLAnow forecasts](#).

Explore the [Google search terms](#).

<https://www.etsa.fi/en/etlanow-eu28/>
Username and password: etlanow2015

ETLAnow Search Terms

Belgium (BE)

Un

ETLAnow: Google search terms for Belgium

Close

1. werkloosheidsverzekering
2. werkloosheidstoelage
3. werkloosheidsuitkering
4. l'indemnisation du chômage
5. allocation de chômage
6. secours de chômage
7. assurance chômage
8. Forem
9. VDAB
10. Actiris
11. RVA
12. ONEM
13. Jobat
14. Vacature
15. Stepstone

. = offic
Last up
Next u
Next o
Export
Explor

Bul

Un

Search terms updated: 2015-11-02 09:58:31

Edit

ETLAnow on Twitter

The image shows a screenshot of the ETLAnow Twitter profile. At the top, the profile name 'ETLAnow' is displayed in large blue letters, with 'NOW' in white on a blue square background. Below this, the profile picture is a smaller version of the 'ETLAnow' logo. The bio states: 'ETLAnow is an experiment run by ETLA, The Research Institute of the Finnish Economy to use big data in economic forecasting.' The location is 'Helsinki, Finland' and the website is 'etla.fi/en/etlanow-eu2...'. The statistics show 31 tweets, 11 following, and 2 followers. An 'Edit profile' button is visible. The main content area shows a tweet from ETLAnow @ETLAnow, dated 8 hours ago, predicting 9.07% unemployment for EU on 2015-11-04 based on Google searches, with a 1.13 pp decrease from last year. The tweet includes a link 'ow.ly/S1I9u'. On the right, the 'Who to follow' section lists 'Taloussanomat @talous...' and 'Petri Riikonen @pjriikonen', both with 'Follow' buttons.

ETLAnow 
@ETLAnow

ETLAnow is an experiment run by ETLA, The Research Institute of the Finnish Economy to use big data in economic forecasting.

 Helsinki, Finland
 etla.fi/en/etlanow-eu2...

TWEETS **31** FOLLOWING **11** FOLLOWERS **2** [Edit profile](#)

Tweets Tweets & replies

 **ETLAnow** @ETLAnow  · 8h

ETLAnow predicts 9.07% unemployment for EU on 2015-11-04 based on Google searches; 1.13 pp decrease from last year
ow.ly/S1I9u

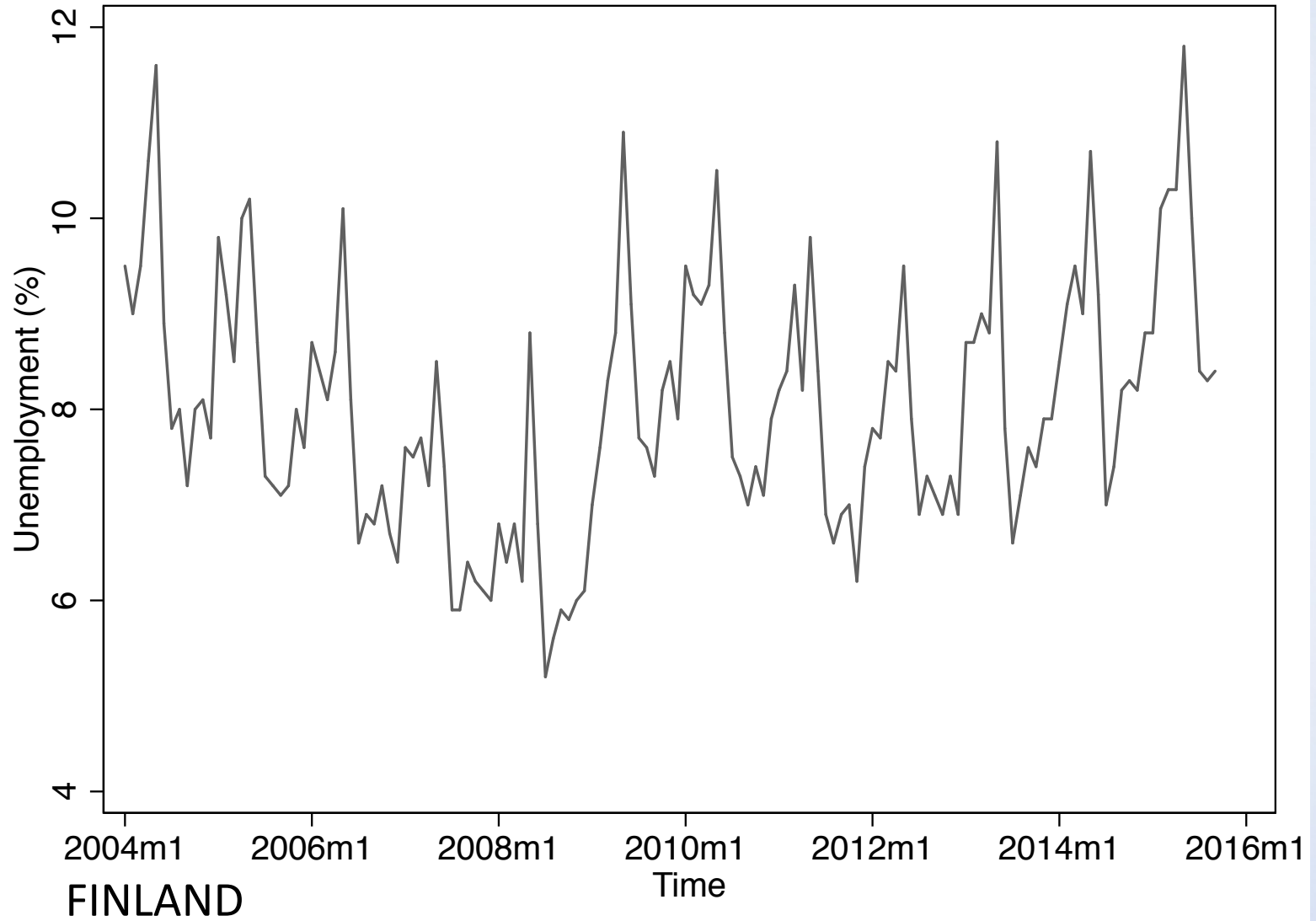
Who to follow · Refresh · View all

 **Taloussanomat**  @talous... 
Followed by Timo Seppala a...
[Follow](#)

 **Petri Riikonen** @pjriikonen 
Followed by Joonas Tuhkuri
[Follow](#)

Find friends

UNEMPLOYMENT

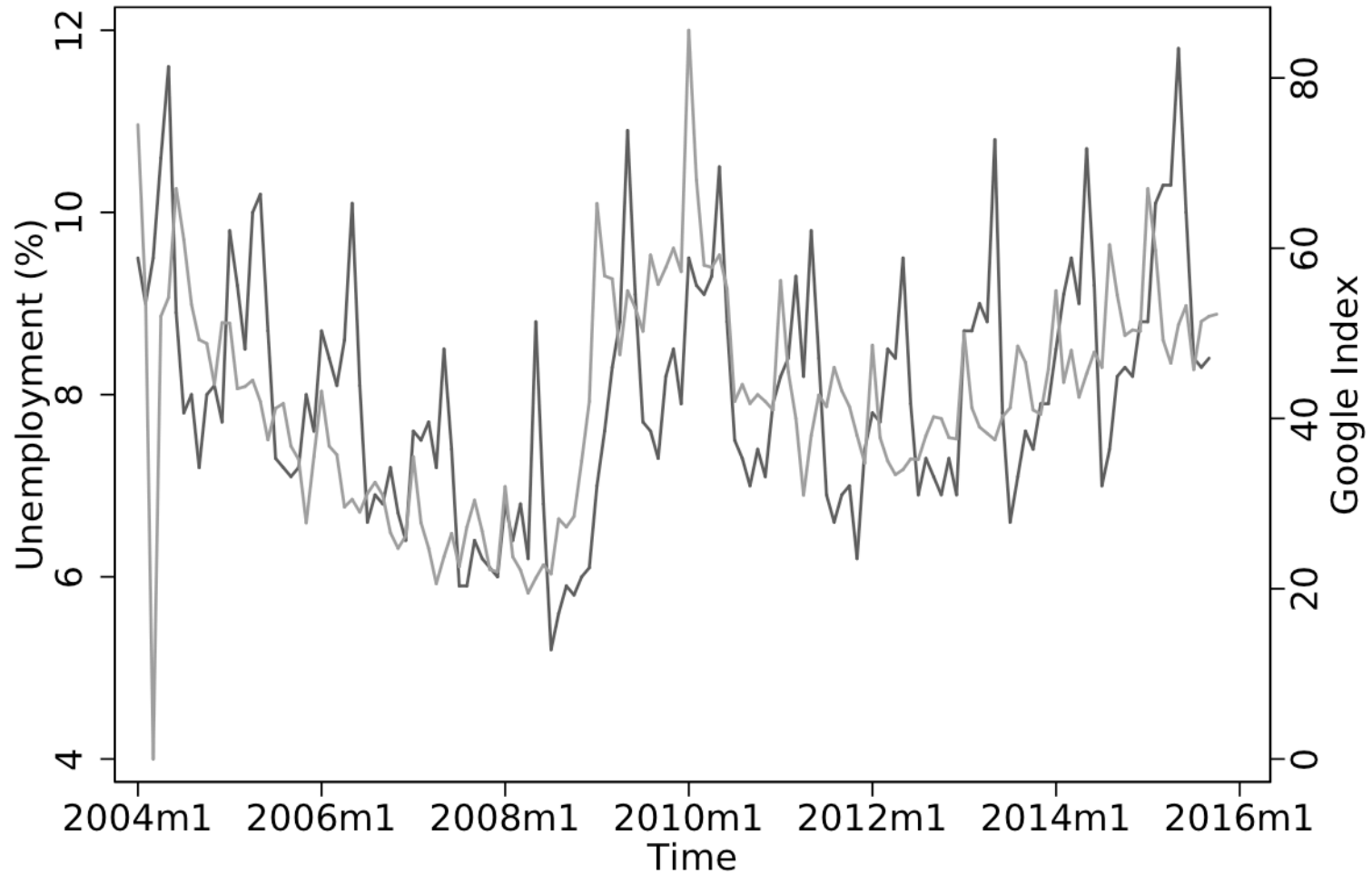


GOOGLE INDEX

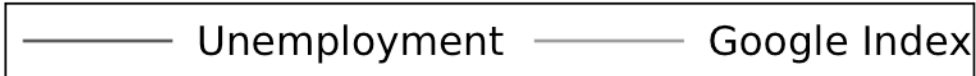
- unemployment benefits
- unemployment office
- unemployment claim
- unemployment compensation
- unemployment insurance*

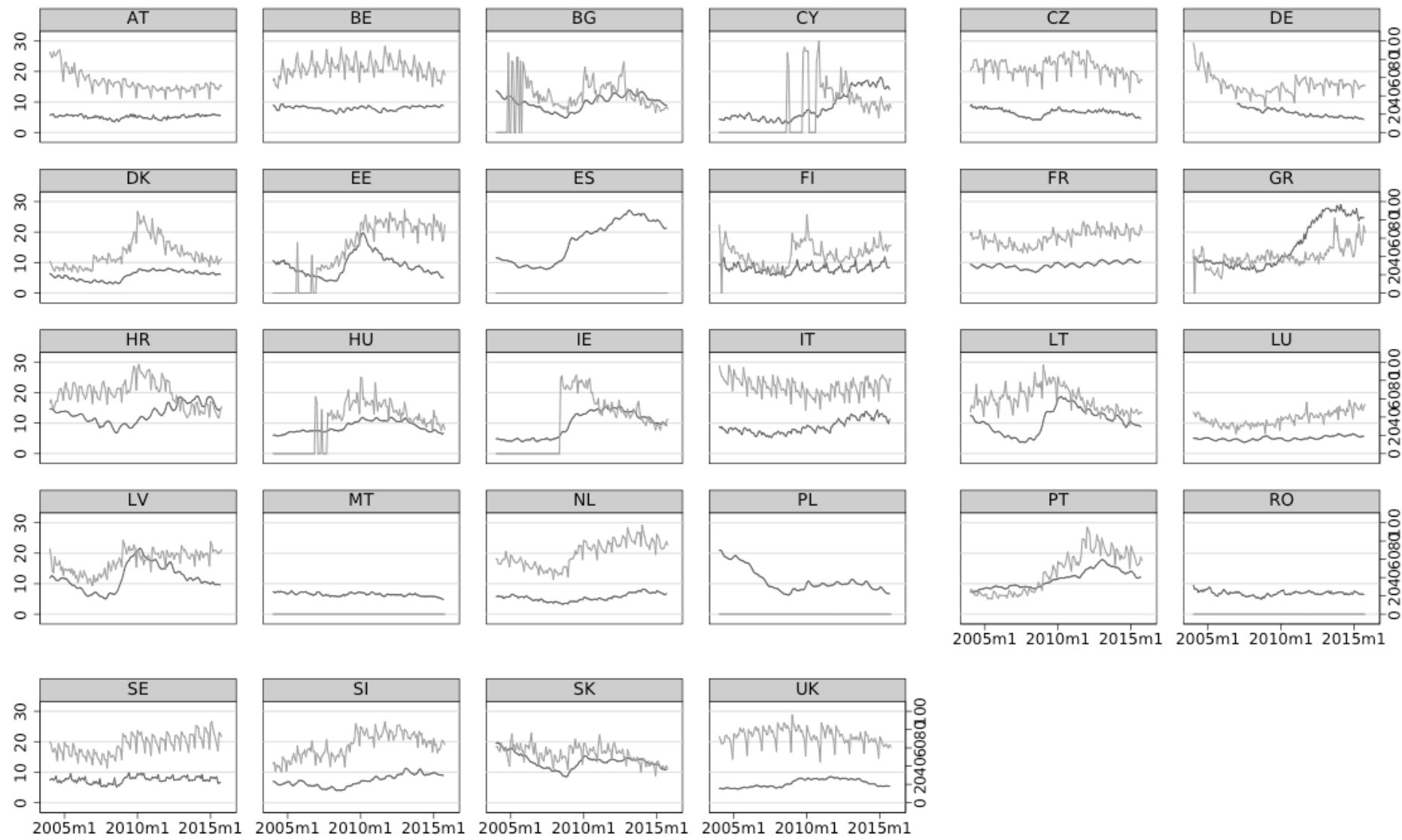
*In Finnish

GOOGLE INDEX



FINLAND

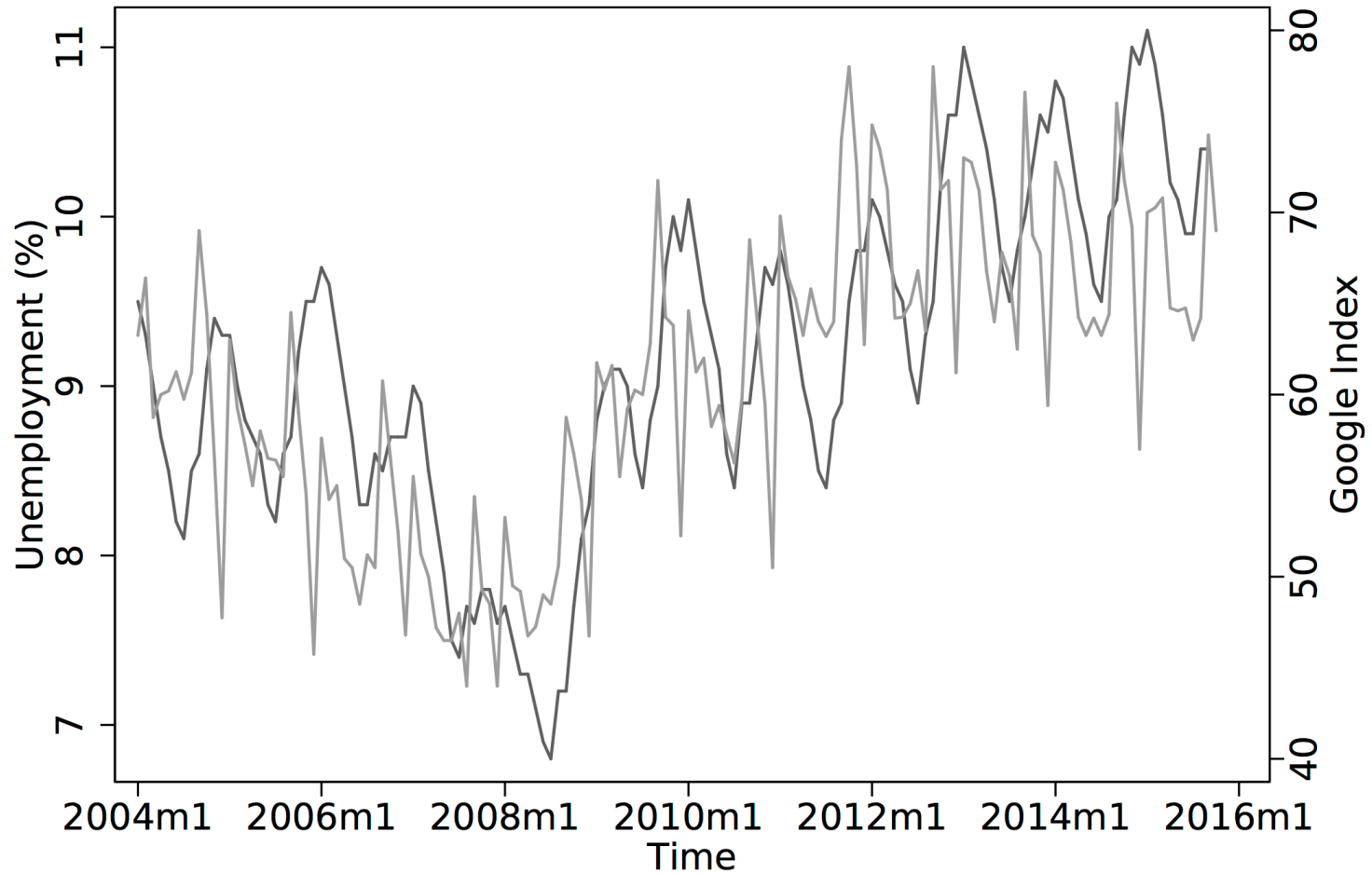




Unemployment
Google Index

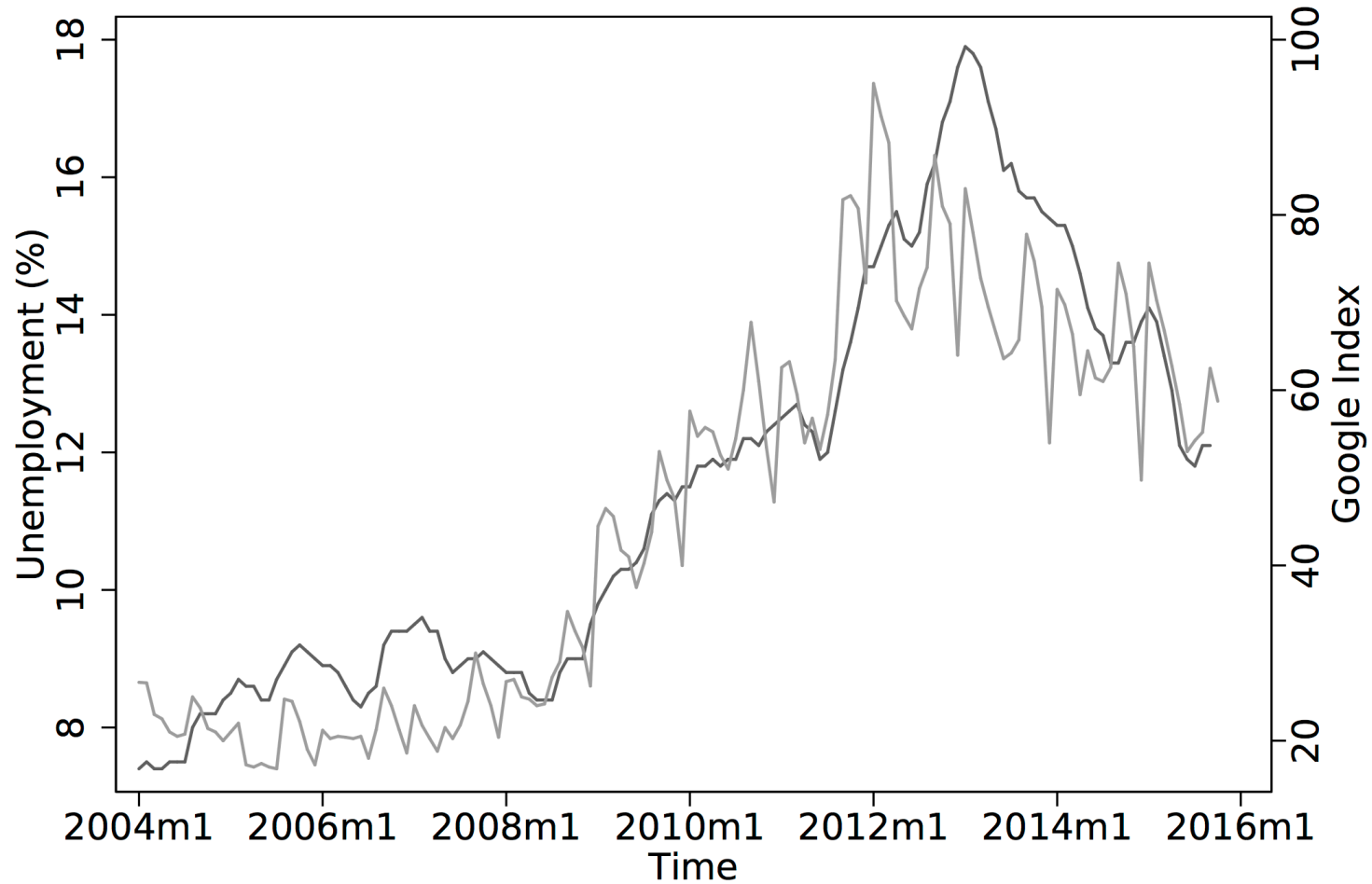
Graphs by Country

FRANCE



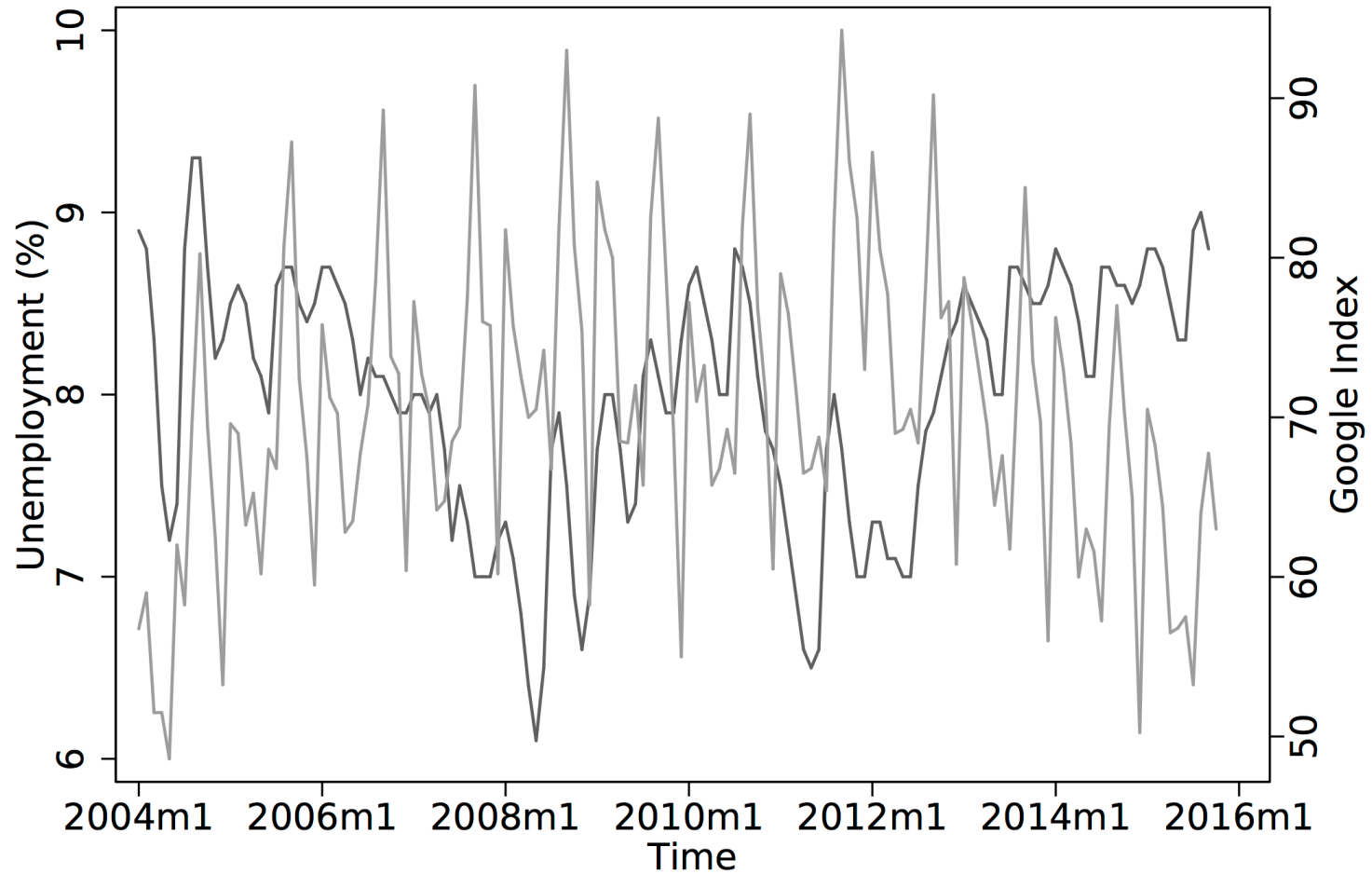
— Unemployment — Google Index

PORTUGAL



— Unemployment — Google Index

BELGIUM



— Unemployment — Google Index

CROSS CORRELATION

CCF

h	-4	-3	-2	-1	0	1	2	3	4
AT	0.17	0.16	0.19	0.28	0.32	0.26	0.27	0.23	0.25
BE	-0.30	-0.38	-0.37	-0.28	-0.17	-0.11	-0.27	-0.50	-0.52
BG	0.12	0.10	0.09	0.06	0.04	0.04	0.03	0.02	0.03
CY	0.47	0.45	0.44	0.41	0.40	0.37	0.36	0.34	0.32
CZ	0.34	0.31	0.30	0.34	0.40	0.39	0.38	0.34	0.27
DE	-0.49	-0.52	-0.53	-0.49	-0.42	-0.43	-0.46	-0.50	-0.51
DK	0.69	0.68	0.68	0.68	0.66	0.63	0.58	0.55	0.53
EE	0.37	0.38	0.38	0.39	0.39	0.40	0.40	0.41	0.40
FI	0.64	0.55	0.45	0.42	0.45	0.36	0.32	0.38	0.31
FR	0.72	0.73	0.74	0.72	0.70	0.63	0.57	0.54	0.52

$n = 130$, $h =$ lag of Google Index, CCF = value of cross-correlation function. The values of CCF on the left-hand side tell the correlation coefficients between past Google search volumes and the present unemployment.

CCF

h	-4	-3	-2	-1	0	1	2	3	4
GR	0.51	0.53	0.54	0.55	0.58	0.59	0.58	0.59	0.59
HR	-0.41	-0.43	-0.45	-0.48	-0.52	-0.58	-0.63	-0.66	-0.68
HU	0.72	0.74	0.75	0.76	0.77	0.74	0.71	0.68	0.66
IE	0.83	0.82	0.81	0.81	0.79	0.76	0.73	0.70	0.67
IT	-0.08	-0.10	-0.03	0.05	0.14	0.01	-0.01	0.02	0.10
LT	0.23	0.19	0.15	0.12	0.08	0.05	0.02	-0.01	-0.05
LU	0.59	0.59	0.58	0.57	0.58	0.60	0.60	0.61	0.62
LV	0.64	0.64	0.63	0.62	0.60	0.58	0.56	0.54	0.51
NL	0.75	0.73	0.72	0.71	0.72	0.64	0.60	0.57	0.56
PL	0.89	0.90	0.91	0.92	0.92	0.91	0.87	0.87	0.85
SE	0.56	0.57	0.56	0.52	0.52	0.35	0.45	0.40	0.36
SI	0.66	0.62	0.61	0.60	0.58	0.55	0.51	0.48	0.45
SK	0.30	0.30	0.31	0.34	0.37	0.34	0.30	0.27	0.26
UK	0.10	0.06	0.05	0.08	0.08	0.02	-0.05	-0.10	-0.11

$n = 130$, $h =$ lag of Google Index, CCF = value of cross-correlation function. The values of CCF on the left-hand side tell the correlation coefficients between past Google search volumes and the present unemployment.

GRANGER CAUSALITY

Null hypothesis

Country	VAR(1)				VAR(1) using lead of x			
	$y \rightarrow x$		$x \rightarrow y$		$y \rightarrow x$		$x \rightarrow y$	
	χ^2	p -value	χ^2	p -value	χ^2	p -value	χ^2	p -value
AT	<0.001	0.99	0.65	0.42	1.78	0.18	3.56	0.059
BE	0.20	0.66	11.3	0.001***	2.20	0.14	10.6	0.001***
BG	0.089	0.77	4.02	0.045*	0.049	0.82	3.93	0.047*
CY	1.81	0.18	5.06	0.024*	1.97	0.16	4.61	0.032*
CZ	9.18	0.002**	0.32	0.57	9.85	0.002**	1.98	0.16
DE	9.25	0.002**	25.0	<0.001***	11.4	0.001**	0.11	0.74
DK	0.94	0.33	6.86	0.009**	0.20	0.66	15.2	<0.001**

y = unemployment rate, x = Google Index.

The sample period is Jan 2004–Oct 2014 ($n = 130$). Both models estimated are first-order VARs, which, based on the Schwarz criterion, are statistically adequate simplifications of second-order VARs. Asterisks * and ** denote significance at the 5% and 1% levels, i.e., Granger non-causality ' \rightarrow ' is rejected.

Null hypothesis

Country	VAR(1)				VAR(1) using lead of x			
	$y \rightarrow x$		$x \rightarrow y$		$y \rightarrow x$		$x \rightarrow y$	
	χ^2	p -value	χ^2	p -value	χ^2	p -value	χ^2	p -value
EE	1.50	0.22	0.12	0.73	0.94	0.33	0.014	0.91
FI	0.78	0.38	4.12	0.043*	1.31	0.25	11.2	0.001**
FR	11.1	0.001**	11.6	0.001**	7.17	0.007**	26.8	0.000**
GR	15.5	<0.001**	0.80	0.37	15.9	<0.001**	0.41	0.52
HR	14.0	<0.001**	5.75	0.016**	16.5	<0.001**	15.8	<0.001**
HU	11.9	0.001**	1.60	0.21	8.77	0.003**	5.49	0.019*
IE	0.37	0.54	33.0	<0.001**	0.20	0.65	42.7	<0.001**
IT	0.30	0.58	11.2	0.001**	0.01	0.93	35.4	<0.001**
LT	0.011	0.92	17.8	<0.001**	0.08	0.76	10.49	0.001**
LU	15.2	<0.001**	1.10	0.29	13.7	<0.001**	.082	0.78
LV	7.30	0.007**	9.77	0.002**	5.53	0.019*	8.01	0.005**
NL	0.19	0.67	2.44	0.12	1.23	0.27	58.9	<0.001**
PL	9.35	0.002**	15.7	<0.001**	4.87	0.027*	26.7	<0.001**
SE	0.03	0.86	9.37	0.002**	14.0	<0.001**	28.8	<0.001
SI	3.62	0.057	4.33	0.037*	1.99	0.158	10.13	0.001**
SK	3.83	0.05*	0.41	0.52	2.73	0.098	4.58	0.032**
UK	0.01	0.92	0.015	0.900	0.52	0.468	40.50	<0.001**

MODEL



MODEL

- Fit the best model you can using the data you have (which may often be past values of the time series itself.)
- Add Google Trends data as an additional predictor
- See how the *out-of-sample* forecast improves using mean absolute error using a rolling window forecast.
- Particularly interest in turning points since they are the hardest thing to forecast.

*Choi, Hyunyoung, and Hal Varian. "Predicting the present with google trends." Economic Record 88.1 (2012): 2-9

MODEL

$$\log(y_t) = \beta_0 + \beta_1 \log(y_{t-1}) + \beta_2 \log(y_{t-12}) + e_t$$

$$\log(y_t) = \beta_{00} + \beta_{10} \log(y_{t-1}) + \beta_{20} \log(y_{t-12}) + \beta_{30} x_t + e_t$$



Unemployment rate



Lag



Seasonal
effects



Google Index

Country		h=0	h=1	h=2	h=3	h=4
AT	(0)	6.30	7.45	8.21	8.38	8.73
	(1)	6.37	7.37	8.04	8.33	8.65
	Δ	-1.13	+1.15	+1.97	+0.52	+0.84
BE	(0)	3.11	5.37	6.82	6.93	6.31
	(1)	3.09	5.44	6.47	6.65	6.27
	Δ	+0.45	-1.20	+5.18	+4.09	+0.73
BG	(0)	3.53	6.88	9.94	12.7	14.6
	(1)	3.67	6.96	10.0	12.7	14.7
	Δ	-3.92	-1.12	-0.90	-0.50	-0.54
CY	(0)	6.48	12.9	10.7	14.4	16.2
	(1)	6.46	13.0	10.8	14.3	15.7
	Δ	+0.36	-0.92	-0.48	+1.02	+2.86
CZ	(0)	5.71	7.36	8.66	10.23	10.86
	(1)	5.73	7.31	8.69	10.43	10.82
	Δ	-0.19	+0.72	-0.34	+1.95	+0.38
DE	(0)	4.74	6.45	7.98	9.18	9.21
	(1)	4.29	6.54	7.49	8.92	9.16
	Δ	+9.44	-1.43	+6.19	+2.82	+0.51

(0) and (1) = mean absolute percentage errors (MAPE)

Δ = improvement in forecasting accuracy (neg. % change)

Country		h=0	h=1	h=2	h=3	h=4
EE	(0)	4.65	8.14	11.44	14.17	16.72
	(1)	4.66	8.17	11.49	14.05	16.98
	Δ	-0.27	-0.39	-0.45	+0.82	-1.55
FI	(0)	7.53	9.16	10.1	10.61	10.71
	(1)	6.33	7.03	6.64	6.87	7.61
	Δ	+15.87	+23.23	+34.22	+35.27	+28.93
FR	(0)	2.65	4.56	5.90	6.95	7.68
	(1)	2.57	4.59	5.91	6.92	7.59
	Δ	+2.91	-0.56	-0.08	+0.42	+1.15
GR	(0)	4.97	6.85	8.53	10.9	12.3
	(1)	5.08	6.98	8.85	11.0	13.0
	Δ	-2.24	-1.85	-3.81	-0.22	-5.76
HR	(0)	3.95	7.63	10.95	13.71	15.56
	(1)	3.90	7.57	10.75	13.74	15.68
	Δ	+1.23	+0.85	+1.87	-0.21	-0.76
HU	(0)	2.54	4.57	6.54	8.23	9.60
	(1)	2.52	4.50	6.54	8.31	9.58
	Δ	+0.87	+1.45	+0.07	-0.96	-0.22

(0) and (1) = mean absolute percentage errors (MAPE)

Δ = improvement in forecasting accuracy (neg. % change)

Country		h=0	h=1	h=2	h=3	h=4
IE	(0)	2.95	5.45	7.77	9.80	11.9
	(1)	2.99	5.69	7.94	9.85	12.0
	Δ	-1.45	-4.40	-2.18	-0.54	-1.01
IT	(0)	5.84	8.16	9.59	10.5	11.5
	(1)	4.90	8.27	9.61	10.5	11.7
	Δ	+16.1	-1.36	-0.21	+0.62	-1.73
LT	(0)	4.09	7.50	10.71	13.6	15.9
	(1)	4.20	7.54	10.72	13.7	16.1
	Δ	-2.64	-0.46	-0.14	-0.78	-1.12
LU	(0)	2.87	4.93	6.69	7.96	8.84
	(1)	2.96	4.94	6.67	8.02	8.83
	Δ	-3.21	-0.05	+0.19	-0.86	+0.09
LV	(0)	3.32	6.21	8.48	10.14	11.6
	(1)	3.38	6.21	8.60	10.19	11.6
	Δ	-1.80	+0.15	-1.40	+0.43	+0.36
NL	(0)	3.52	5.75	7.60	8.93	10.26
	(1)	2.84	5.88	7.65	9.29	10.26
	Δ	+19.29	+2.30	-0.84	-4.06	+4.18

(0) and (1) = mean absolute percentage errors (MAPE)

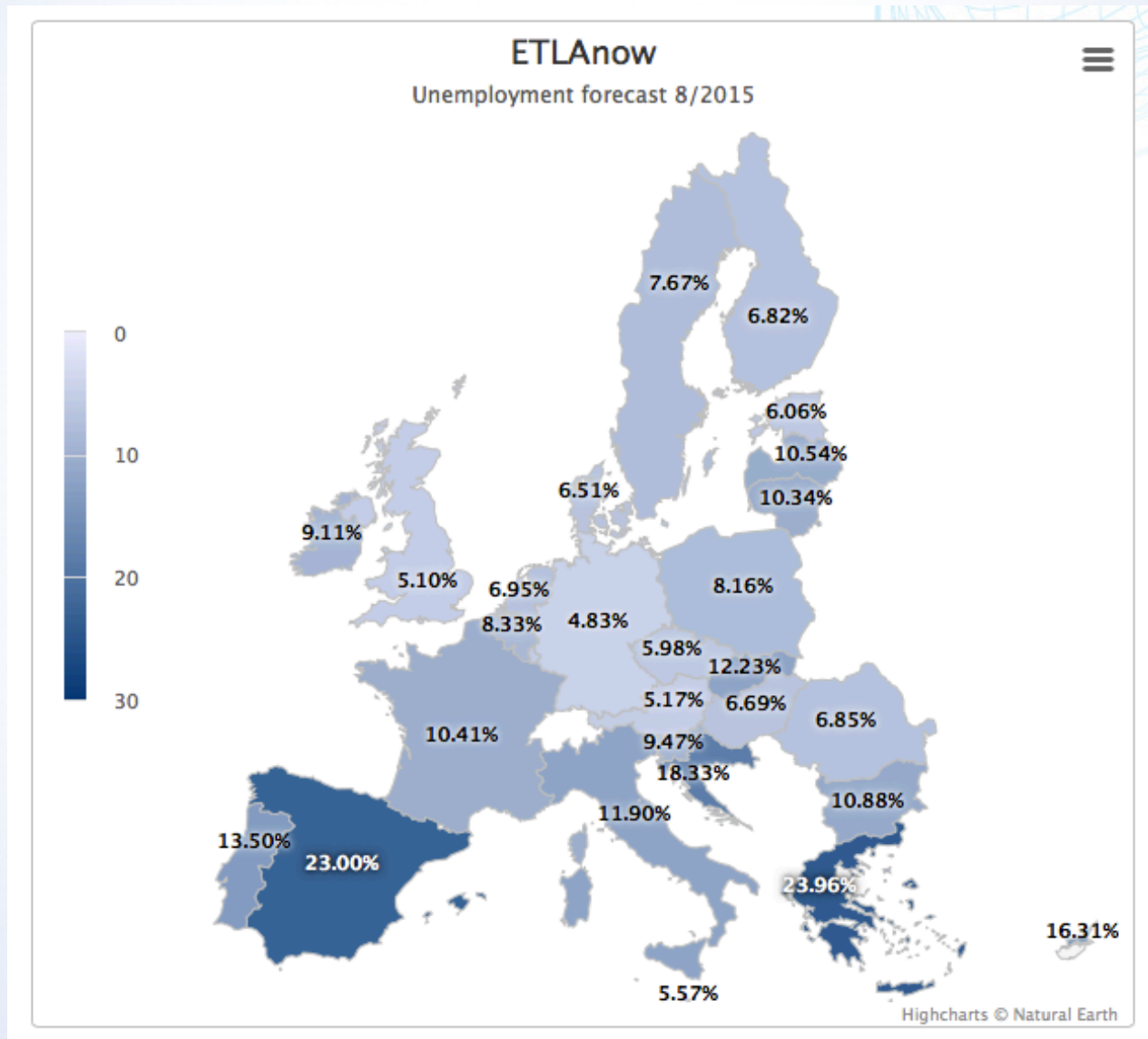
Δ = improvement in forecasting accuracy (neg. % change)

Country		h=0	h=1	h=2	h=3	h=4
PT	(0)	1.79	3.37	4.81	6.19	7.43
	(1)	1.84	3.22	4.84	6.25	7.63
	Δ	-2.73	+4.49	-0.62	-0.84	-2.71
SE	(0)	6.07	7.52	8.50	9.24	9.45
	(1)	5.69	6.05	5.92	5.64	5.81
	Δ	+6.21	+19.57	+30.31	+38.90	+38.48
SI	(0)	3.25	5.62	7.94	9.90	11.41
	(1)	3.31	5.63	8.00	10.14	11.05
	Δ	+1.74	-0.18	-0.66	-2.41	+3.11
SK	(0)	1.92	3.39	4.82	6.49	8.12
	(1)	1.88	3.41	4.81	6.42	8.12
	Δ	+2.04	-0.53	+0.039	+1.06	-0.01
UK	(0)	2.30	4.14	5.42	6.35	7.49
	(1)	2.15	4.08	5.39	6.36	7.33
	Δ	+6.91	+1.37	+0.46	-0.06	+2.10

(0) and (1) = mean absolute percentage errors (MAPE)

Δ = improvement in forecasting accuracy (neg. % change)

PANEL DATA



PANEL DATA

Model	FE (AB)	FE (OLS)
<hr/>		
Variables		
<hr/>		
$\log(y_{t-1})$	0.959** (0.00580)	0.958** (0.00545)
$\log(y_{t-12})$	0.00889 (0.00550)	0.00891* (0.00515)
x_t	0.000745** (0.000072)	0.000763** (0.000068)
<hr/>		
Summary statistics for FE (OLS)		
<hr/>		
R^2	within	0.963
	between	0.997
	overall	0.980
F test that state fixed effects = 0	4.67 (<0.0001)	

VARIABLES

- No improvements using search volumes for *Facebook*
- Results vary between countries
- Possible solution: better search terms

CONCLUSION

- Google searches predict unemployment
- Limited to short-term predictions
- Value for forecasting purposes episodic
- Improvements still small
- But useful for economic forecasting

ETLA NOW

<https://www.etla.fi/en/etlanow-eu28/>
Username and password: etlanow2015