BIG DATA: GOOGLE SEARCHES PREDICT UNEMPLOYMENT

4TH BALTIC-NORDIC CONFERENCE ON SURVEY STATISTICS

BaNoCoSS 26 AUG 2015,

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

WORKING PAPER: GOOGLE SEARCHES PREDICT UNEMPLOYMENT:

HOW FAR, WHEN, AND HOW MUCH?

ETLA

UNIVERSITY OF HELSINKI

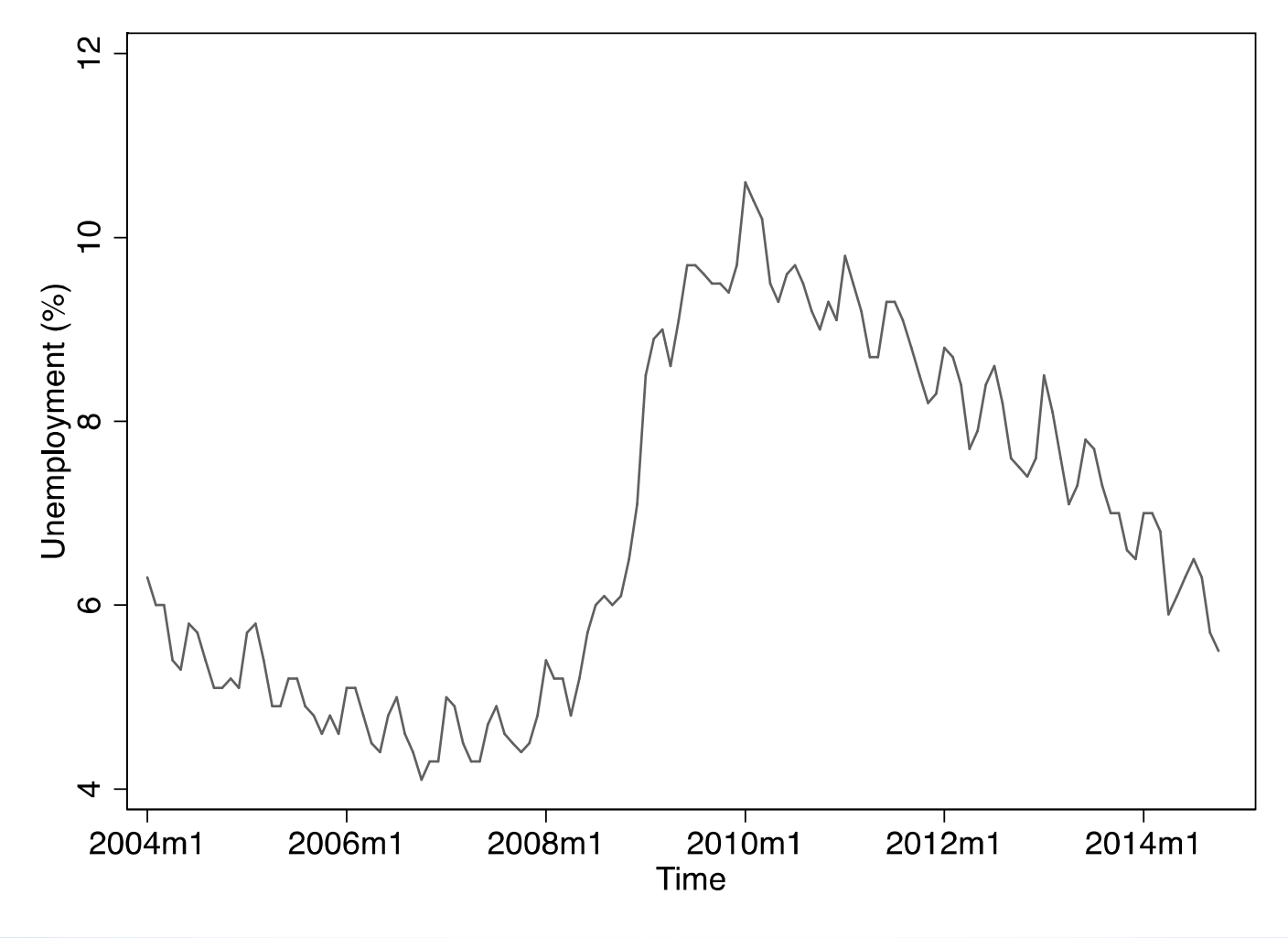
100 B

LITERATURE

- What we know: Google searches predict unemployment.[1]
- What we don’t know: How far, when, and how much?

[1]Askitas & Zimmermann (2009), Choi and Varian (2012), Tuhkuri (2014)

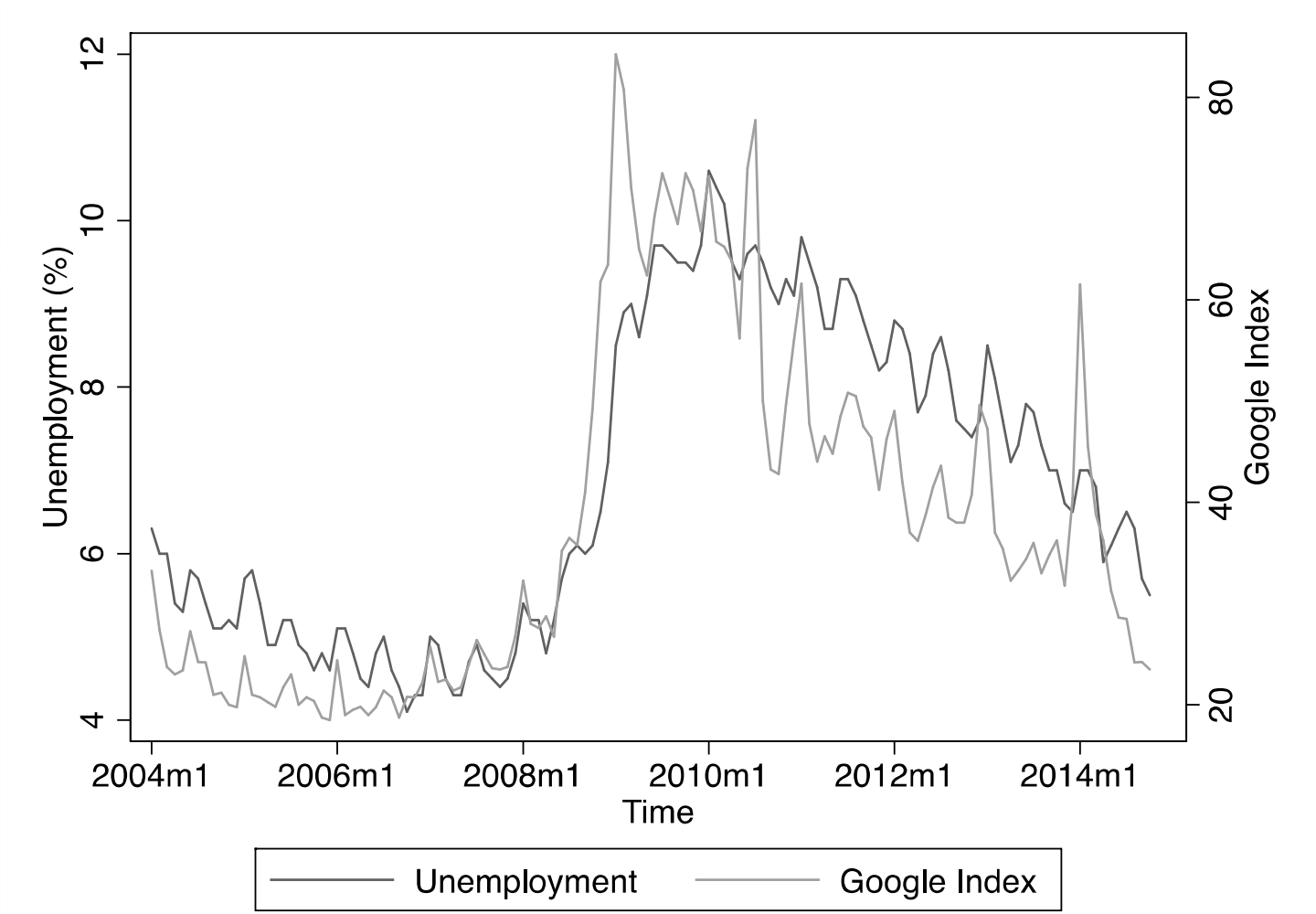
UNEMPLOYMENT



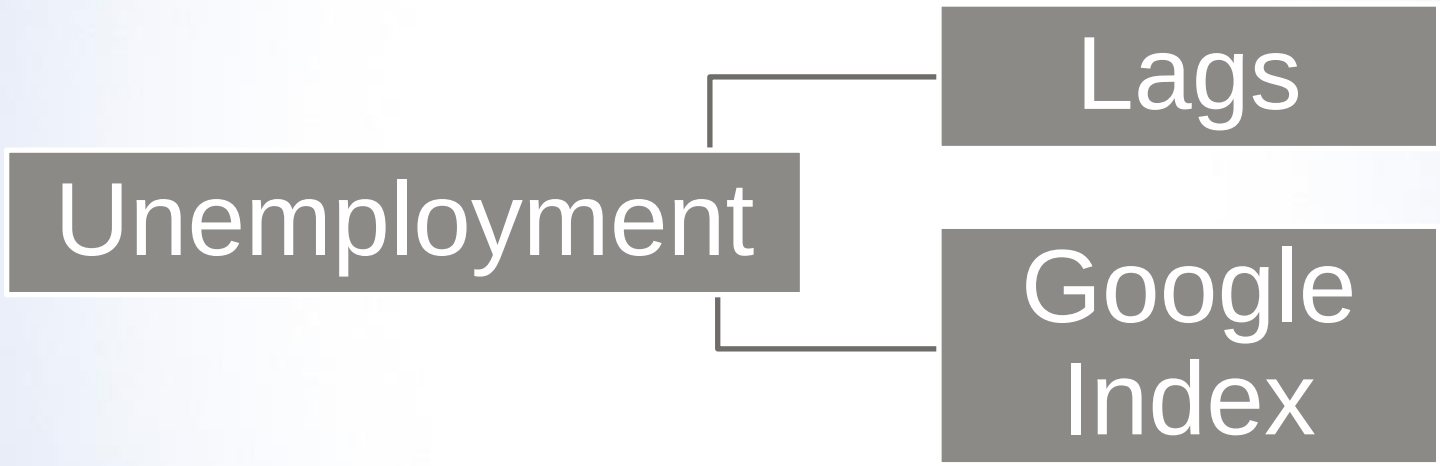
GOOGLE INDEX

- unemployment benefits
- unemployment office
- unemployment claim
- unemployment compensation
- unemployment insurance
- apply for unemployment
- applying for unemployment
- filing for unemployment
- unemployment online
- unemployment office locations
- unemployment eligibility
- ui benefits
- unemployment benefit

GOOGLE INDEX



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.

[1]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

CROSS-CORRELATION

h	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
CCF	.92	.91	.89	.88	.89	.89	.87	.82	.77	.74	.72	.70	.67

h = lag of Google Index,
CCF = value of cross correlation function

GRANGER CAUSALITY

Null hypothesis

VAR(1)				VAR(1) using lead of x			
$y \nrightarrow x$		$x \nrightarrow y$		$y \nrightarrow x$		$x \nrightarrow y$	
χ^2	p-value	χ^2	p-value	χ^2	p-value	χ^2	p-value
0.040	0.84	22.83	$<$0.001***	0.0032	0.96	71.6	$<$0.001***

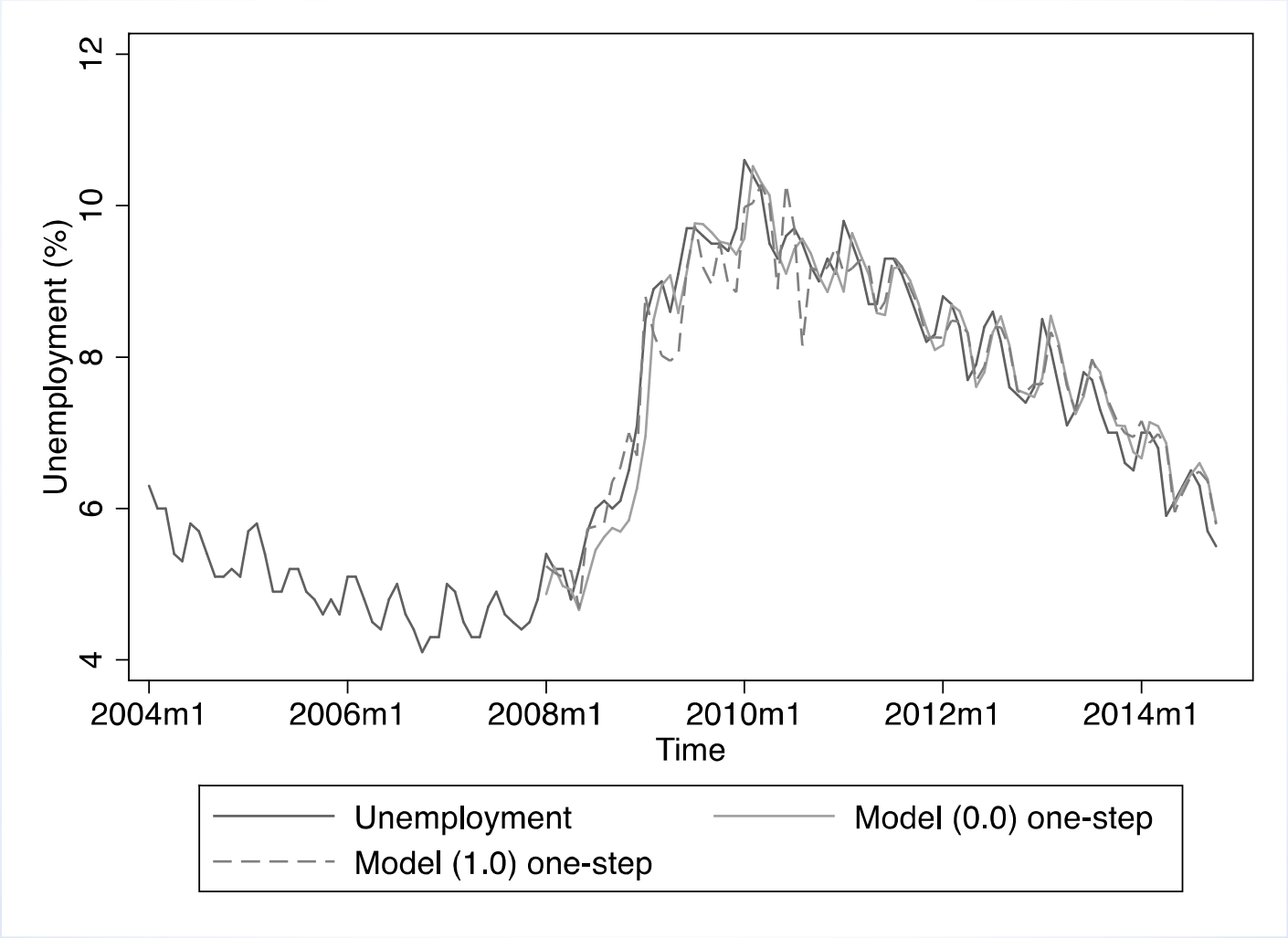
y = unemployment rate, x = Google Index.

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%, 1% and 0.1% levels, i.e. Granger non-causality $' \nrightarrow '$ is rejected.

MODEL

Model	(0.0)	(1.0)
Variables		
$log(y_{t-1})$	0.983**	0.955**
	(0.0295)	(0.0356)
$log(y_{t-12})$	-0.0103	0.0156
	(0.0300)	(0.0368)
x_t		.00440**
		(0.000656)
Constant	1.848**	1.692**
	(0.191)	(0.150)
Summary		
R^2	0.962	0.969
AIC	-371.2	-396.6
BIC	-359.7	-382.3

PREDICTING THE PRESENT



PREDICTING THE PRESENT

Model	MAE	Δ
(0.0)	4.58 %	4.32 %
(1.0)	4.38 %	

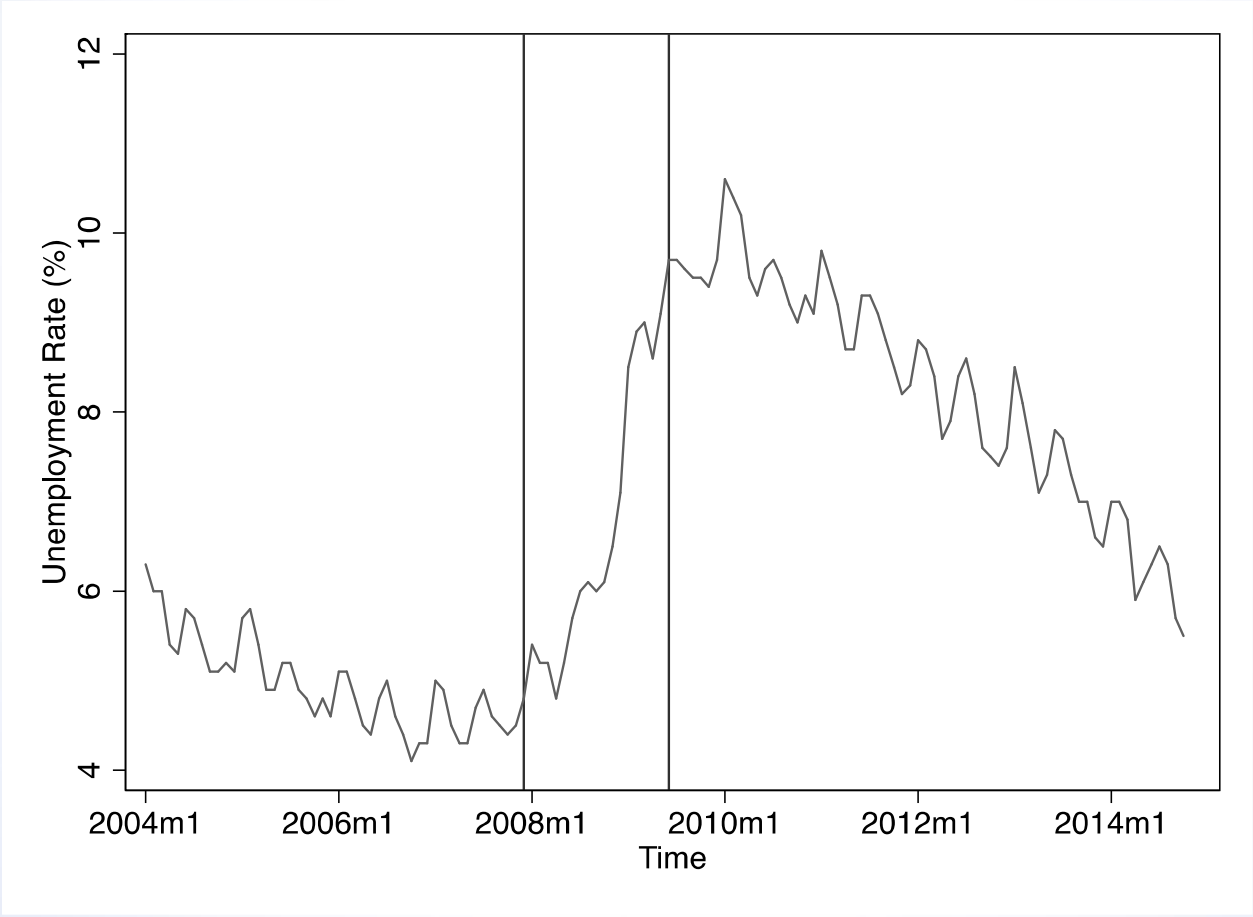
MAE = Mean Absolute Error

Δ = improvement in forecasting accuracy

FORECASTING THE FUTURE

Period	Model	MAE	Δ
t	(0.0) (1.0)	4.58% 4.38%	4.32%
$t+1$	(0.0) (1.1)	7.57% 7.01%	7.48%
$t+2$	(0.0) (1.2)	9.48% 9.85%	-3.92%
$t+3$	(0.0) (1.3)	10.4% 11.06%	-6.28%
$t+4$	(0.0) (1.4)	11.1% 13.02%	-17.22%
$t+5$	(0.0) (1.5)	11.96% 13.54%	-13.22%
$t+6$	(0.0) (1.6)	13.40% 12.07%	9.93%

TIME-SPECIFIC FORECASTS

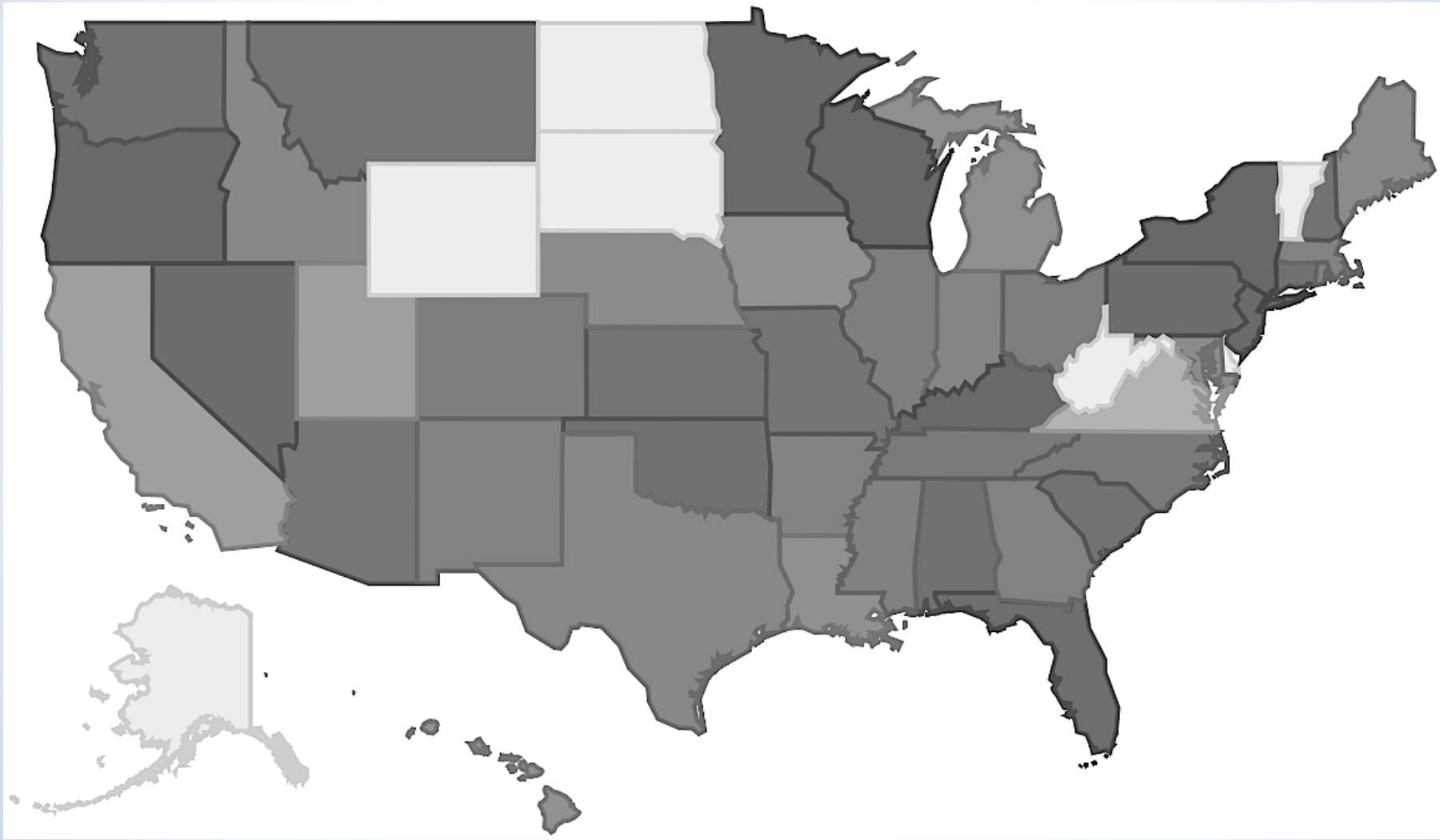


The Recession

TIME-SPECIFIC FORECASTS

Period	Model	MAE	Δ
t	(0.0)	7.17%	17.95%
	(1.0)	5.88%	
$t+1$	(0.0)	11.69%	34.50%
	(1.1)	7.66%	
$t+2$	(0.0)	15.60%	4.53%
	(1.2)	14.89%	
$t+3$	(0.0)	20.57%	-25.57%
	(1.3)	25.57%	
$t+4$	(0.0)	26.07%	-35.06%
	(1.4)	35.06%	

PANEL DATA



PANEL DATA

Model		(1.0)	FE
Variables			
$log(y_{t-1})$		0.955**	0.832**
		(0.0356)	(0.0062)
$log(y_{t-12})$		0.0156	0.0673**
		(0.0368)	(0.00499)
x_t		.00440**	.00167**
		(0.000656)	(0.000066)
Constant		1.692**	
		(0.150)	
Summary statistics for FE			
R^2	within		0.935
	between		0.998
	overall		0.956
F test that state fixed effects = 0			5.51
			(<0.0001)

VARIABLES

- Same results with the most popular search term: unemployment benefit
- No improvements using search volumes for *Facebook*

CONCLUSION

- Google searches predict unemployment
- Limited to short-term predictions
- Value for forecasting purposes episodic
- Improvements smaller than previously reported

BIG DATA: GOOGLE SEARCHES PREDICT UNEMPLOYMENT

4TH BALTIC-NORDIC CONFERENCE ON SURVEY STATISTICS

BaNoCoSS 26 AUG 2015,

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

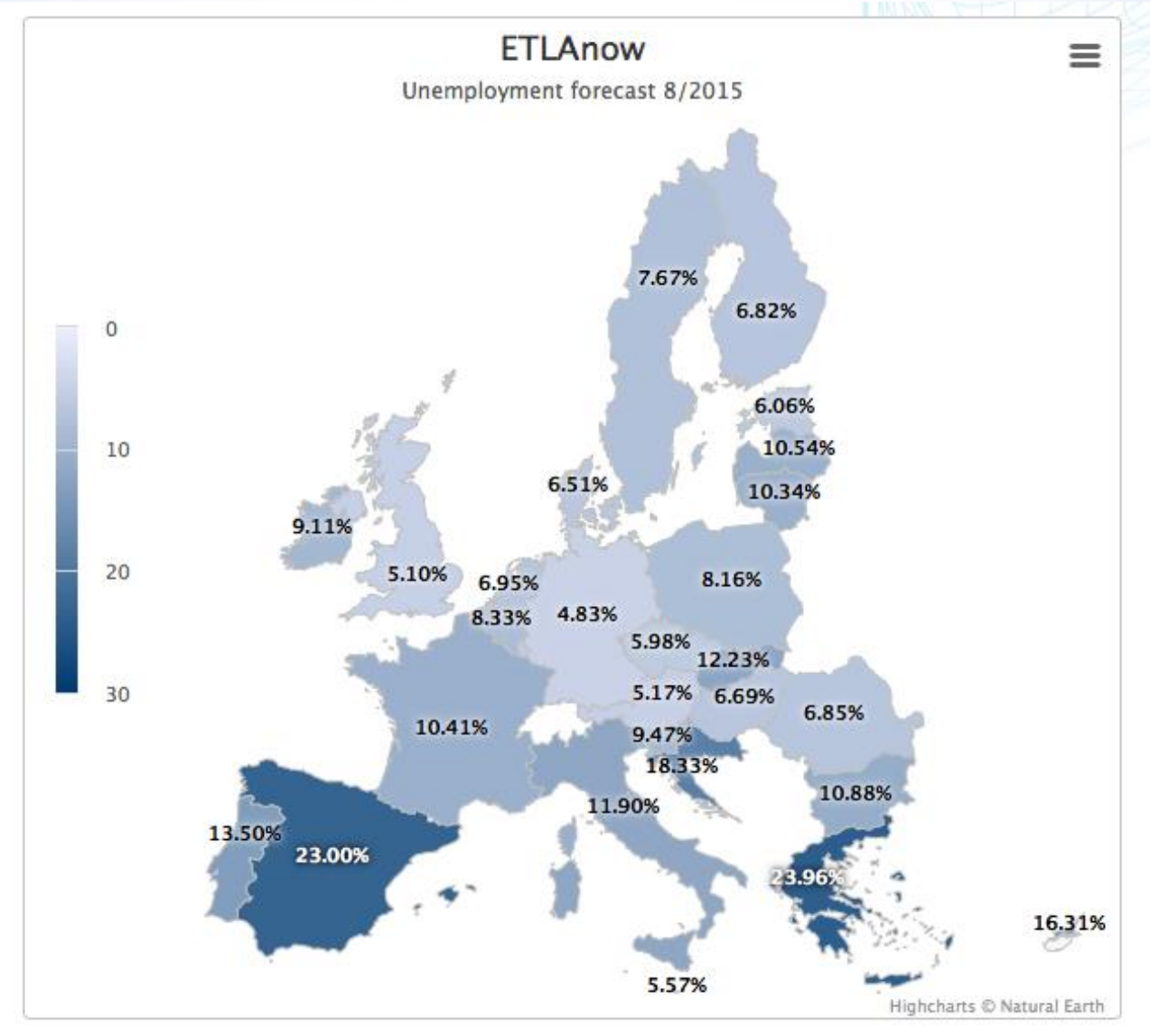
WORKING PAPER: GOOGLE SEARCHES PREDICT UNEMPLYOMENT:

HOW FAR, WHEN, AND HOW MUCH?

ETLA

UNIVERSITY OF HELSINKI

ETLAnow


ETLAnow
Unemployment forecast 8/2015
0
10
20
30
7.67%
6.82%
6.06%
10.54%
10.34%
6.51%
9.11%
5.10%
6.95%
8.33%
4.83%
8.16%
5.98%
12.23%
5.17%
6.69%
6.85%
10.41%
9.47%
18.33%
10.88%
11.90%
13.50%
23.00%
23.96%
16.31%
5.57%
Highcharts © Natural Earth

ETLAnow

EU-28

Unemployment rate (%)	5/2015	6/2015	7/2015	8/2015	9/2015	10/2015	11/2015
Official	9.50	9.30	.	.	.	.	.
ETLAnow	9.67	9.37	9.01	**9.62**	9.53	9.63	9.73
Change (pp)			-0.69	-0.08	-0.37	-0.37	-0.47
							Date: 25.8.2015

. = official data not available
Last update 25.8.2015.
Next update 26.8.2015.
Next official release on the unemployment rate 1.9.2015.
Export ETLAnow forecasts for this table or all ETLAnow forecasts.

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