Migration Forecast EU

June 3, 2022, Berlin

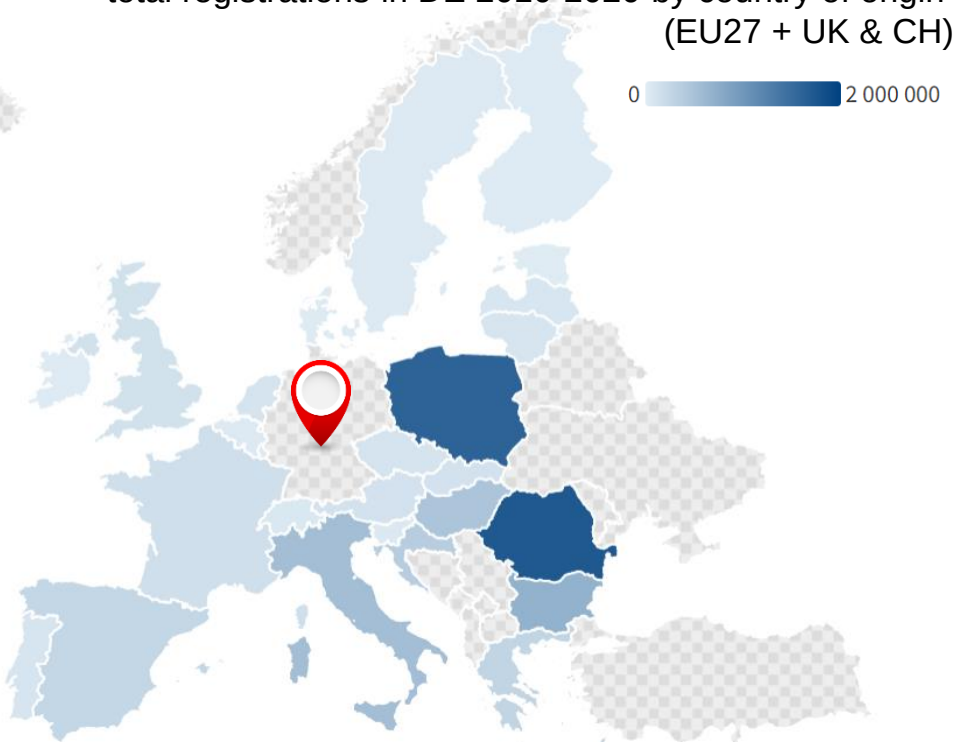
Nicholas Steinbrink and Orkan Kösemen

BertelsmannStiftung

Goal: predict internal EU migration to DE using digital data

- German labour market benefits highly from the **mobility** of EU nationals
- Digital data are supposed to capture **migration intention** and **migration preparation**
- Digital data are **closer to real time** compared to economic data and past migration data
- Does this approach produce better results for short-term forecasts than forecasting with simple time series?

total registrations in DE 2010-2020 by country of origin (EU27 + UK & CH)

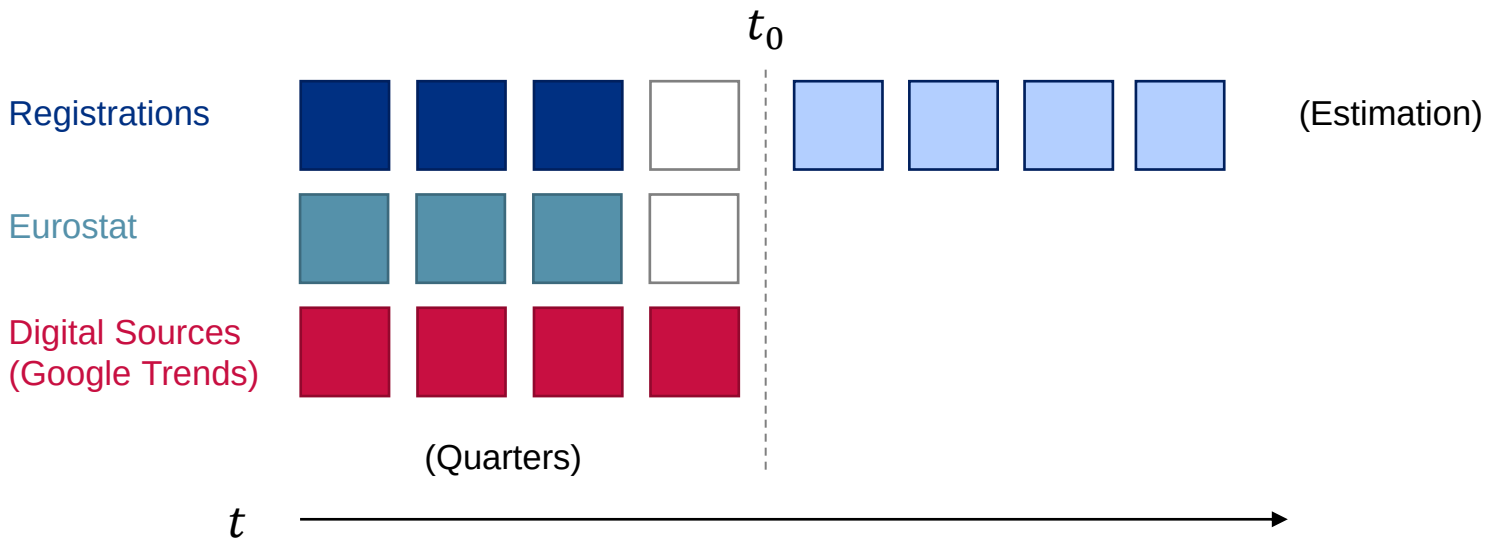


Official data sources

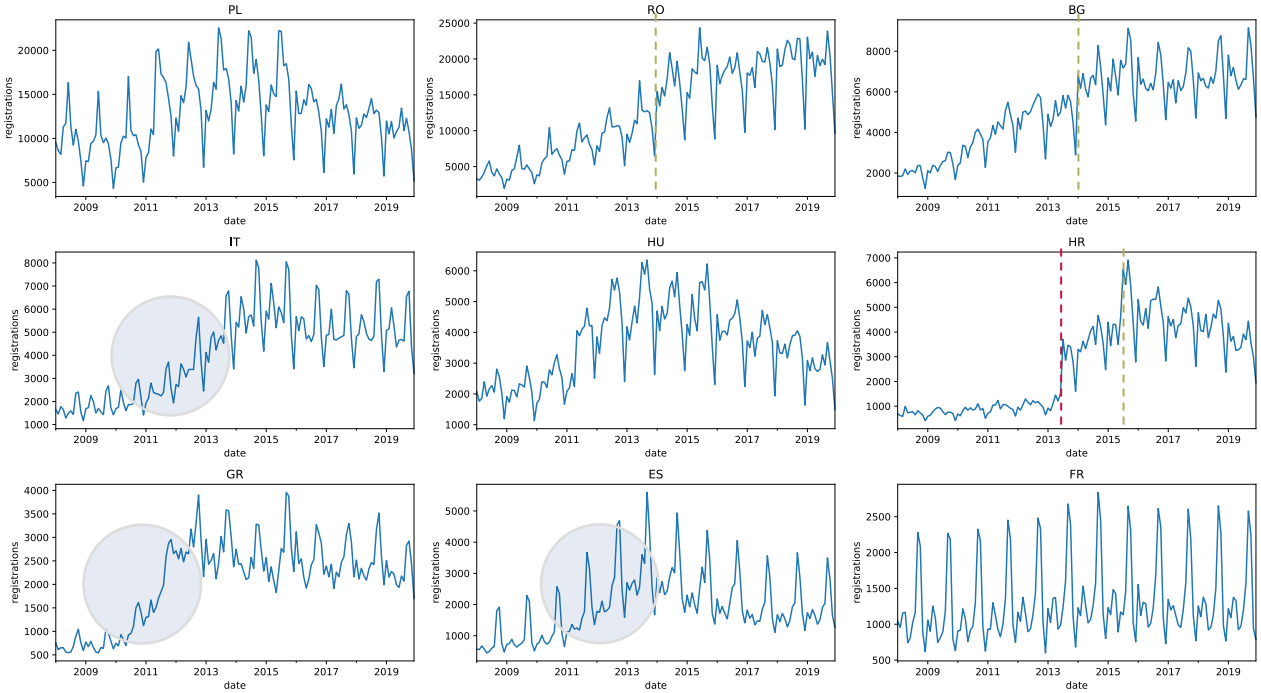
- Prediction target: **registrations** of EU nationals living more than 3 months in DE
  - Availability: monthly
  - Preliminary data available at federal office of statistics (DESTATIS)
  - Latency: 3 – 4 months
  - Desired forecast horizons: 3, 6, (12) months

- Economic and demographic data about EU countries (EUROSTAT)
  - Availability: annually, partly quarterly (e.g. GDP, unemployment)
  - Latency: ~ 1 months (after each quarter)

Forecast setup

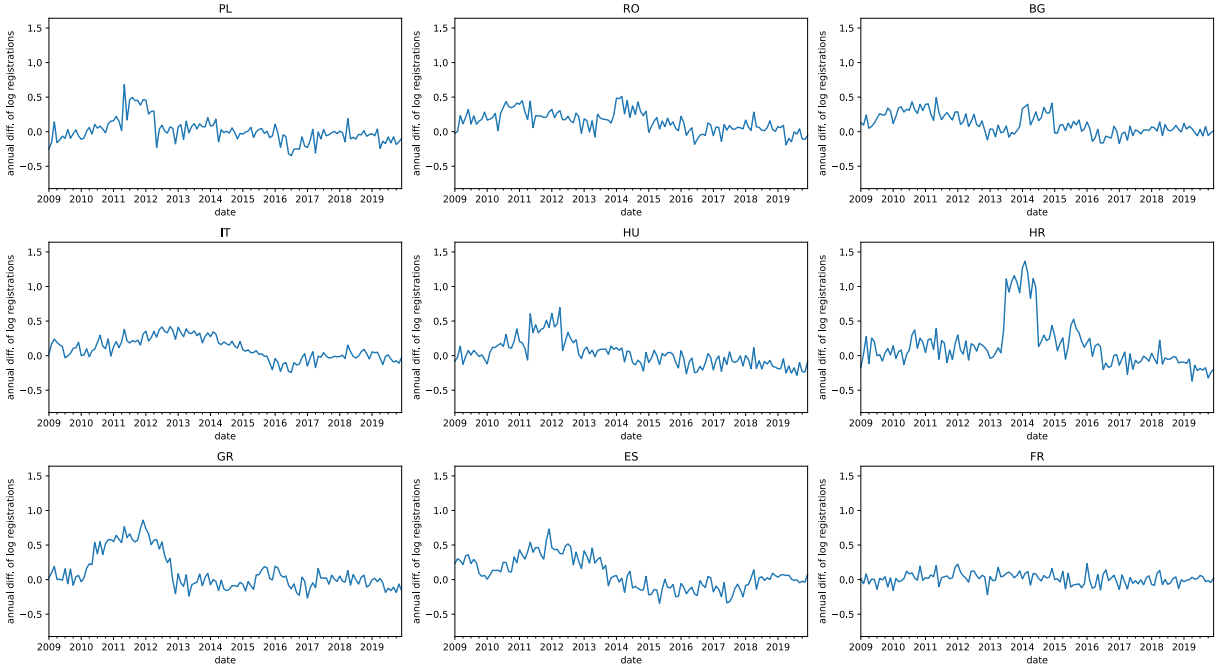


Top 9 countries of origin: vast majority of migration to DE is seasonal



- Periodicity easy to predict but non-seasonal component more challenging
- Main trends:
  - Euro crisis beginning of 2010‘s (especially southern Europe)
  - New EU members and freedom of movement instated (HR, RO, BG)

Prediction target: annual difference of log(registrations)



- We are just interested in the relative change compared to the previous year
- Seasonality disappears almost completely
- Due to transformation, all countries are given equal weight
- Dynamics mostly flat, only few instances of perturbations, therefore harder to predict

Why using Google Trends?

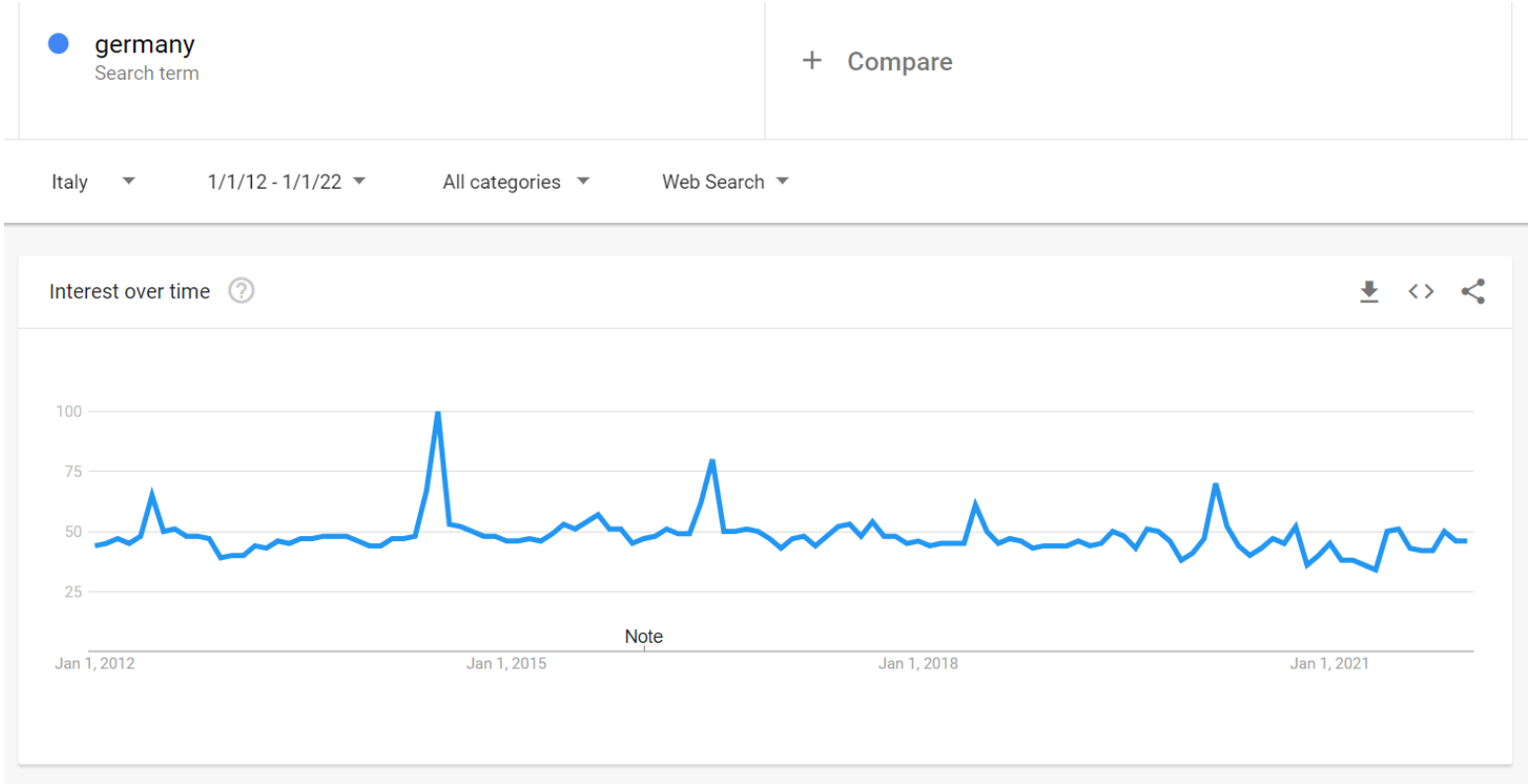
- The correlation between migration flows and migration intention on one side and migration intention and Google searches on the other has been proven, more so for Europe

- Easy public and free access compared to other digital data sources (even though with limitations)

- Most common search tool with wide global usage

- The main idea is to target human behaviour (preferably) after the decision to migrate has been already made in order to remove another layer of uncertainty

Pitfall 1: search terms too unspecific



Pitfall 2: search terms way too specific

health insurance germany
Search term

+ Compare

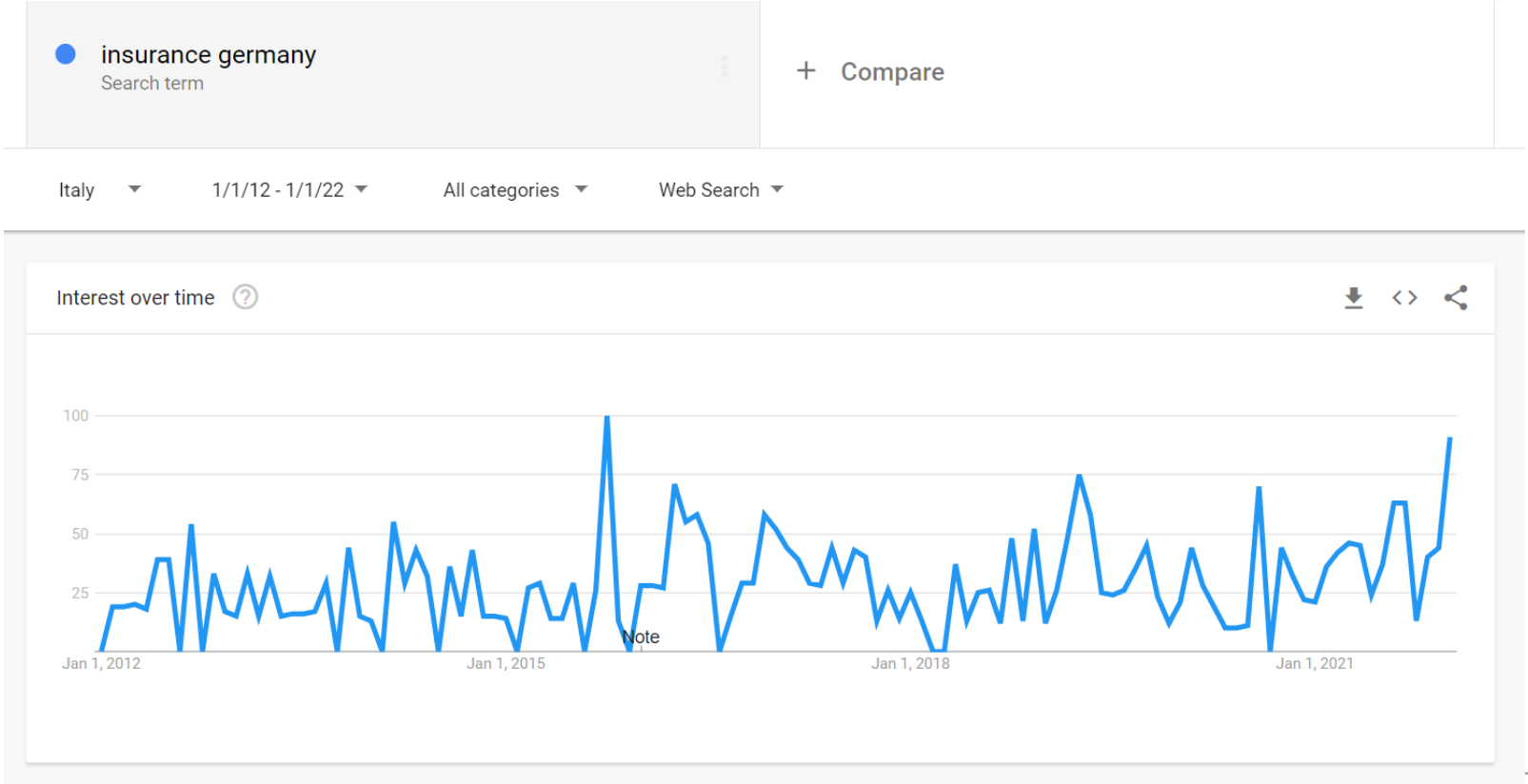
Italy | 1/1/12 - 1/1/22 | All categories | Web Search

Interest over time

Hmm, your search doesn't have enough data to show here.

Please make sure everything is spelled correctly, or try a more general term.

Pitfall 3: search terms slightly too specific

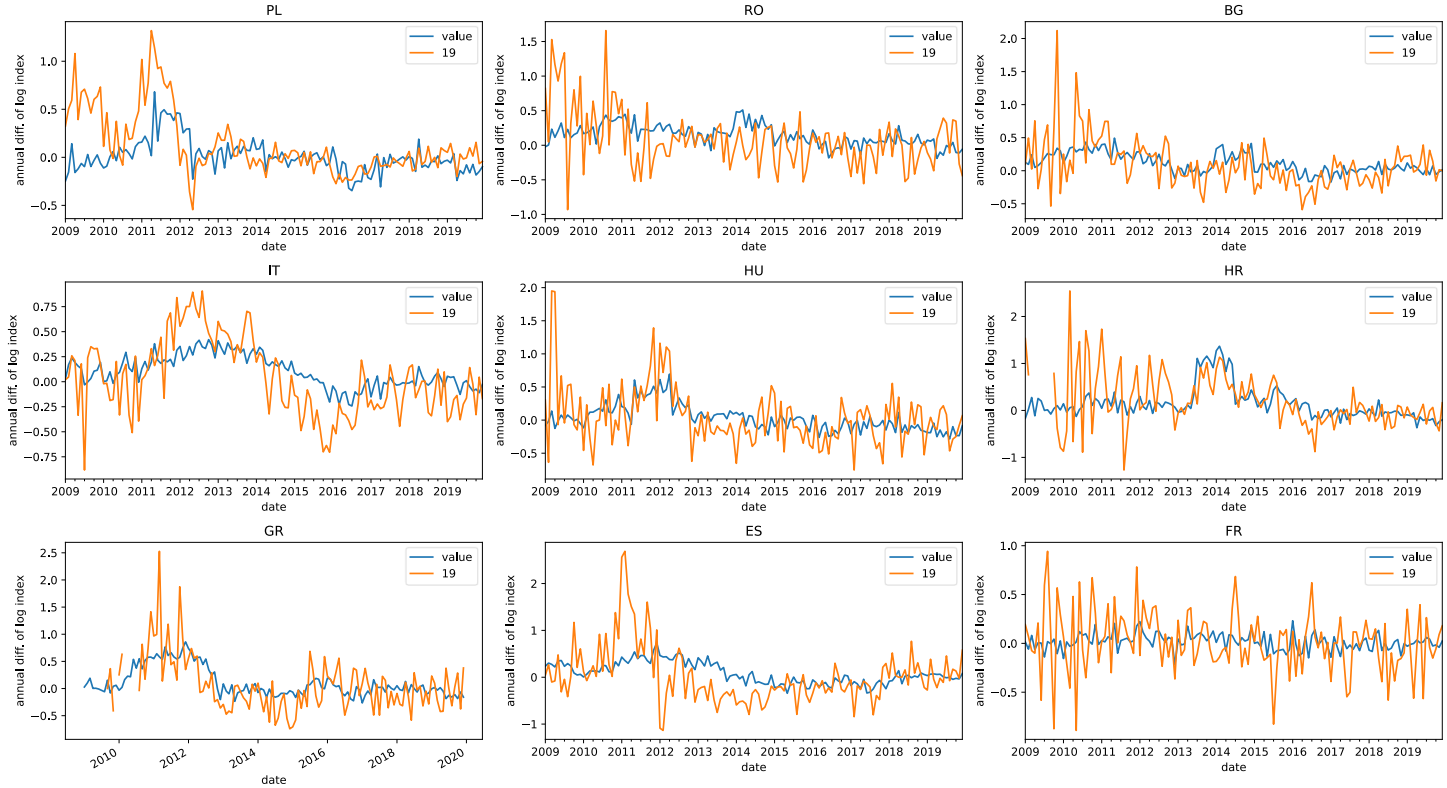


Combining keywords to increase coverage

E.g. keyword group 19: „jobs in germany“

contrato de trabajo alemania + contrato laboral alemania + contrato de empleo alemania + trabajo alemania + empleo alemania + ocupación alemania + ocupacion alemania + trabajar alemania + empleo alemania + empleos alemania + trabajo alemania + trabajos alemania + arbeitsvertrag deutschland + arbeit deutschland + arbeiten deutschland + job deutschland + jobs deutschland + work contract germany + employment germany + working germany + job germany + jobs germany

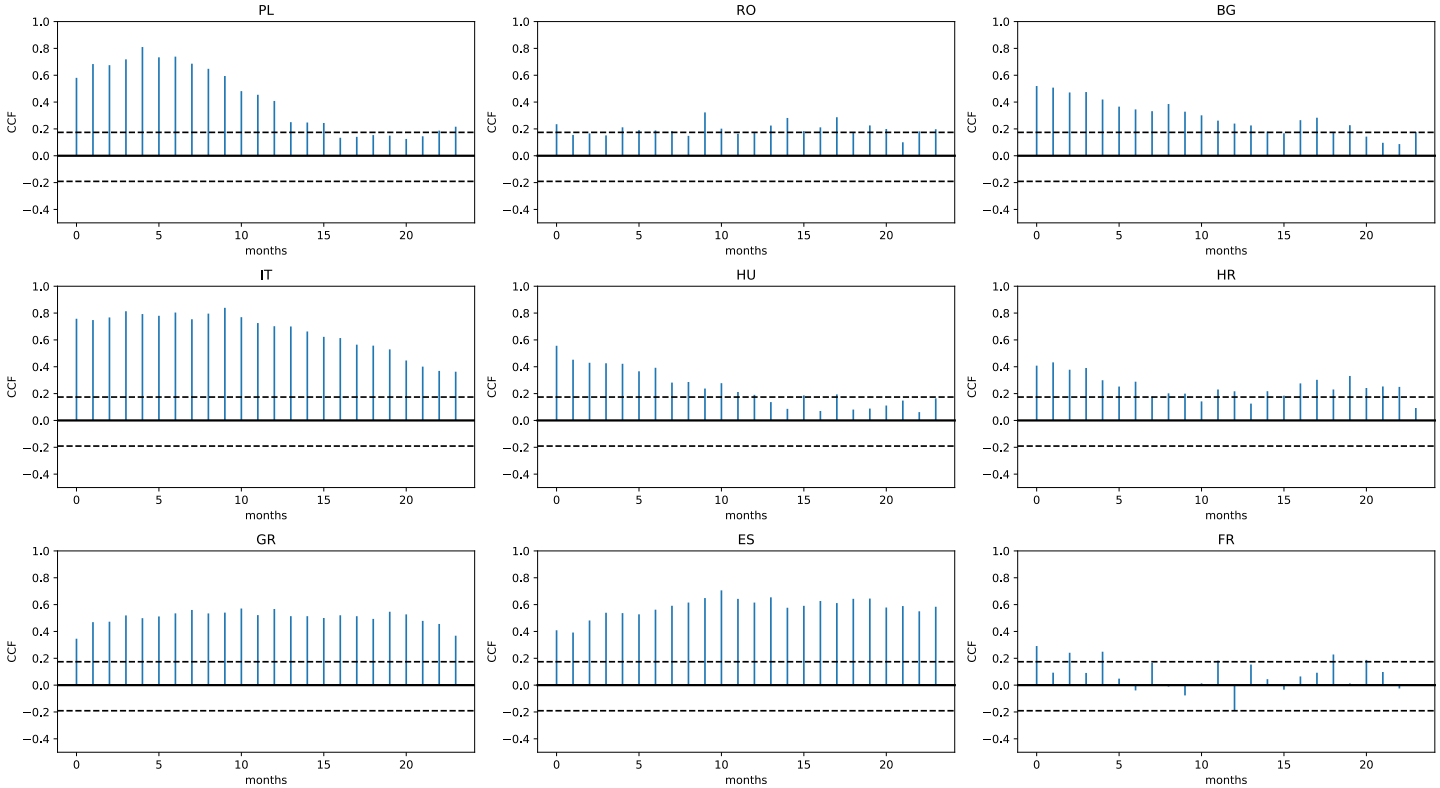
What lag correlation do we have between registrations and Google?



Keyword group 19: jobs/work in germany

(both are transformed)

What lag correlation do we have between registrations and Google?



Range: 2010 - 2019

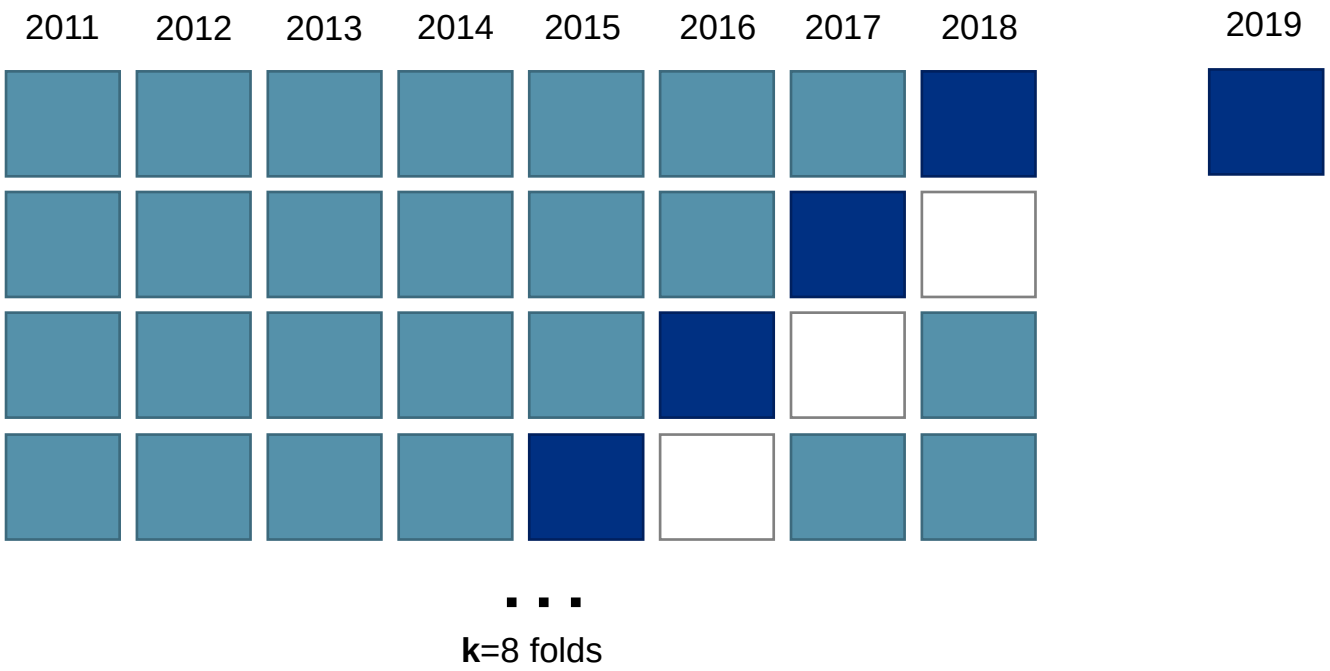
Modeling a panel forecast as a regression problem

t	country	registrations	trends_kw13
01/2022	Transsylvania	**666**	10.0
01/2022	Westeros	**50000**	15.5
02/2022	Transsylvania	**777**	12.0
02/2022	Westeros	**60000**	18.5

t	registrations	registrations_1	registrations_2	trends_kw13_1	transsylvania	westeros
01/2022	**666**	555	444	10.0	1	0
01/2022	**50000**	40000	30000	15.5	0	1
02/2022	**777**	666	555	12.0	1	0
02/2022	**60000**	50000	40000	18.5	0	1

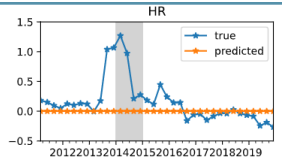
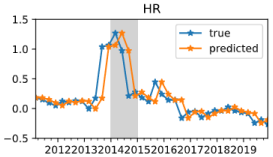
- Can use ML regression algorithms and ML-like feature engineering
- Caveat: Interpretation, forecast uncertainty and n>1 steps ahead not straightforward

Cross Validation Scheme: block k-fold CV + out-of-sample



Alternative to k-fold in presence of correlated residuals
(cf. Bergmeir et al., 2018, https://robjhyndman.com/publications/cv-time-series/)

Benchmarks

Benchmark (transformed)	Benchmark (untransformed)	Example	Comment
Const 0	Same value as previous year		Comparison highlights ability of model to predict phases of high volatility.
Previous value („random walk“)	Same change rate as previous step		Comparison reflects if there is predictive information beyond pure autocorrelation in the model.
Value k steps before („lagged random walk“)	Same change rate as k steps before		More realistic. k should correspond to the delay between data collection and publication (3-4 months)
Best model without Google data	-		Comparison reflects the benefit gained by Google Trends

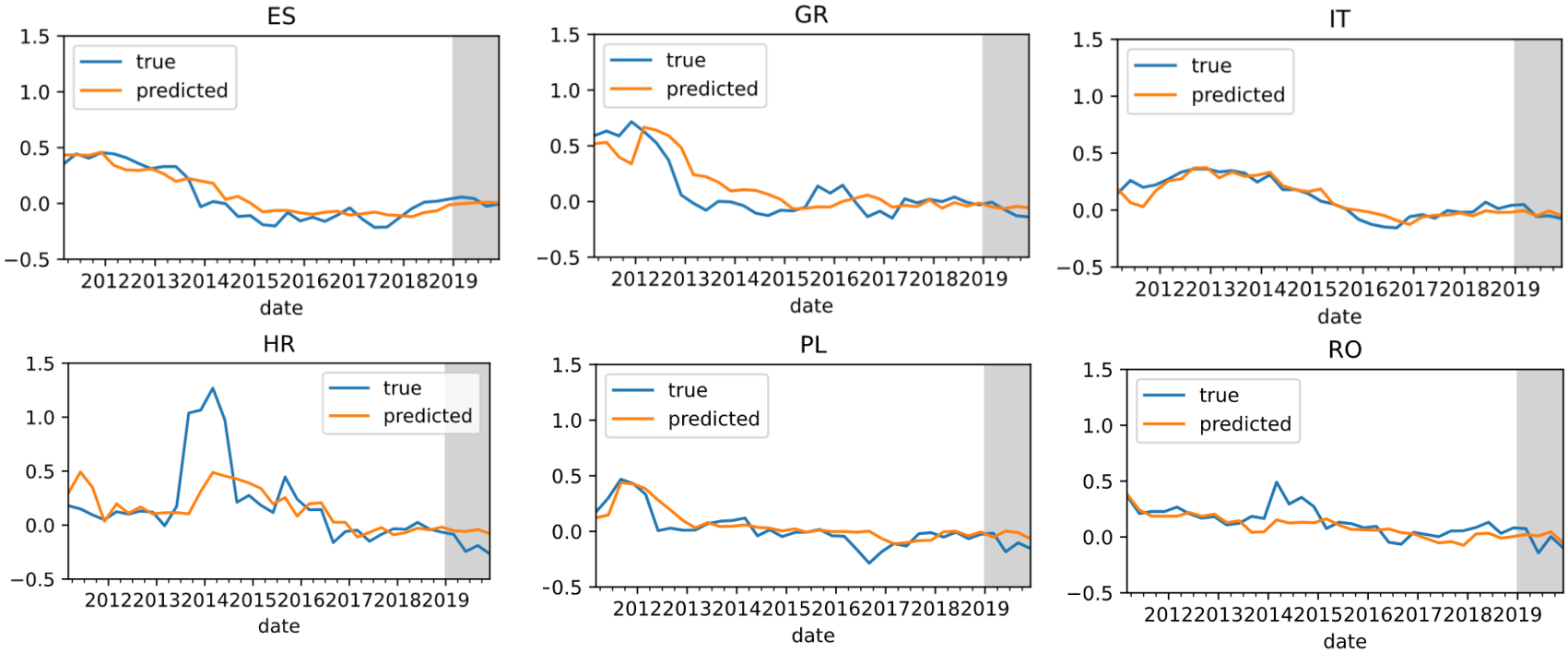
Next quarter forecast: mean CV performance (non-periodic component!)

- Google Trends based model nearly on par with autoregressive one
- Combination minimally beneficial
- However, with real time migration data, naive forecast would be best

Model	Lagged Feature Sets *	R^2_{out}	RMSE	MAE
Benchmark: const. 0	-	0.0	0.127	0.180
Benchmark: previous(1) **	-	**0.71**	**0.105**	**0.070**
Benchmark: previous(2)	-	0.23	0.139	0.092
OLS - autoregressive	Self, GDP, Unemployement	0.37	0.128	0.089
OLS - Trends	Trends Keyword 19, GDP, Unemployment	0.29	0.137	0.095
OLS - combined	Self, Trends Keyword 19, GDP, Unemployment	**0.44**	**0.122**	0.083
Random Forest - autoregressive	Self, Eurostat	0.38	0.129	0.090
Random Forest - Trends	Trends Keyword 19, GDP, Unemployment	0.33	0.134	0.093
Random Forest - combined	Self, Trends Keyword 19, GDP, Unemployment	**0.44**	**0.122**	**0.082**

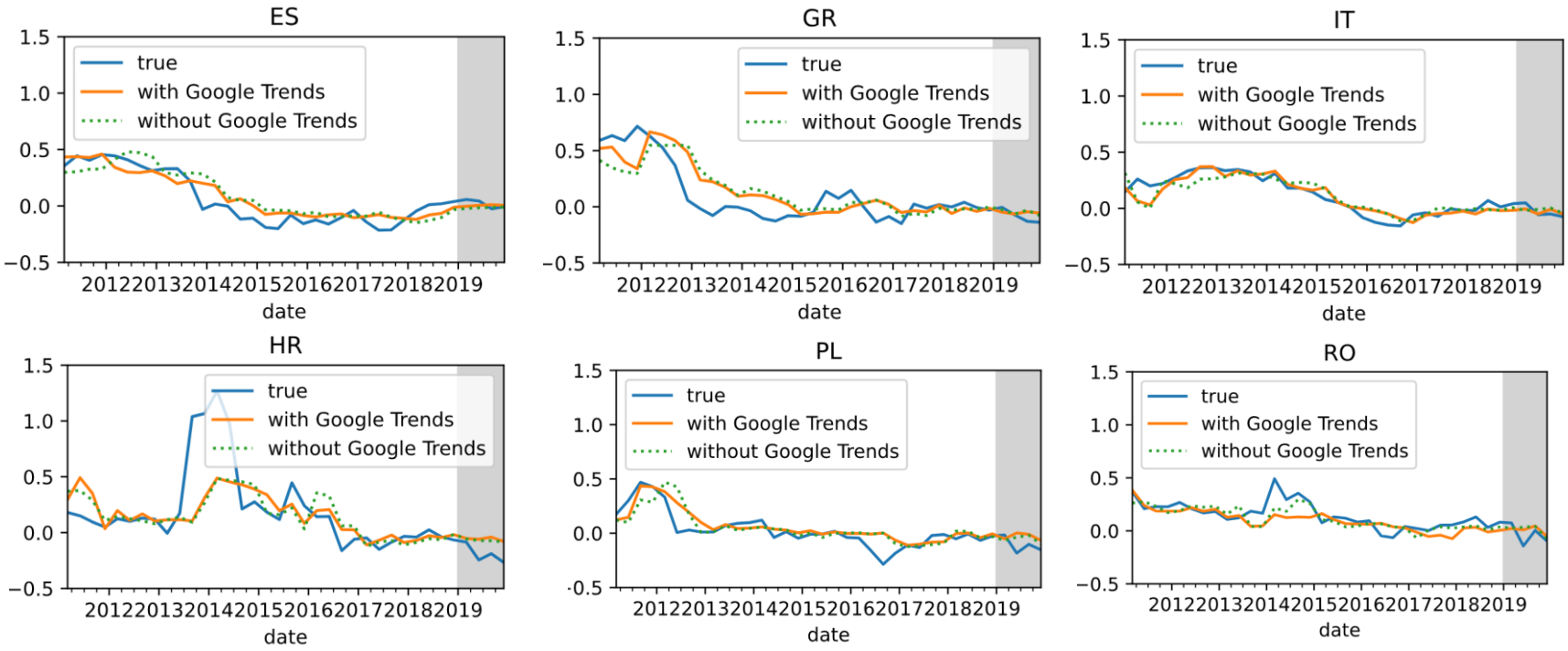
*) all features up to 6 lag periods; self, GDP, unemployment: min. lag of 2 periods
**) not realistic due to publication delay

Random Forest – Trends: some dynamics can be reasonably forecast but not all



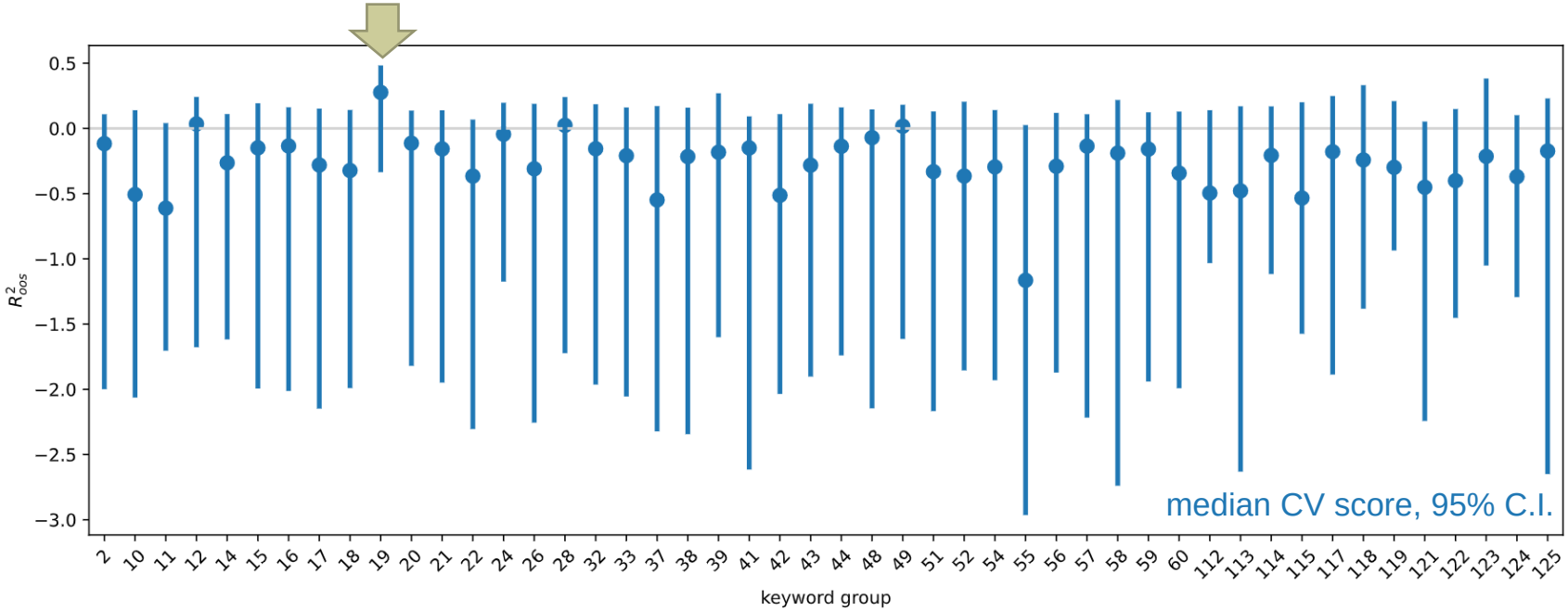
Model: Random Forest
Combine test point predictions from CV; grey area: holdout set

All features combined: partly slight improvements with Google



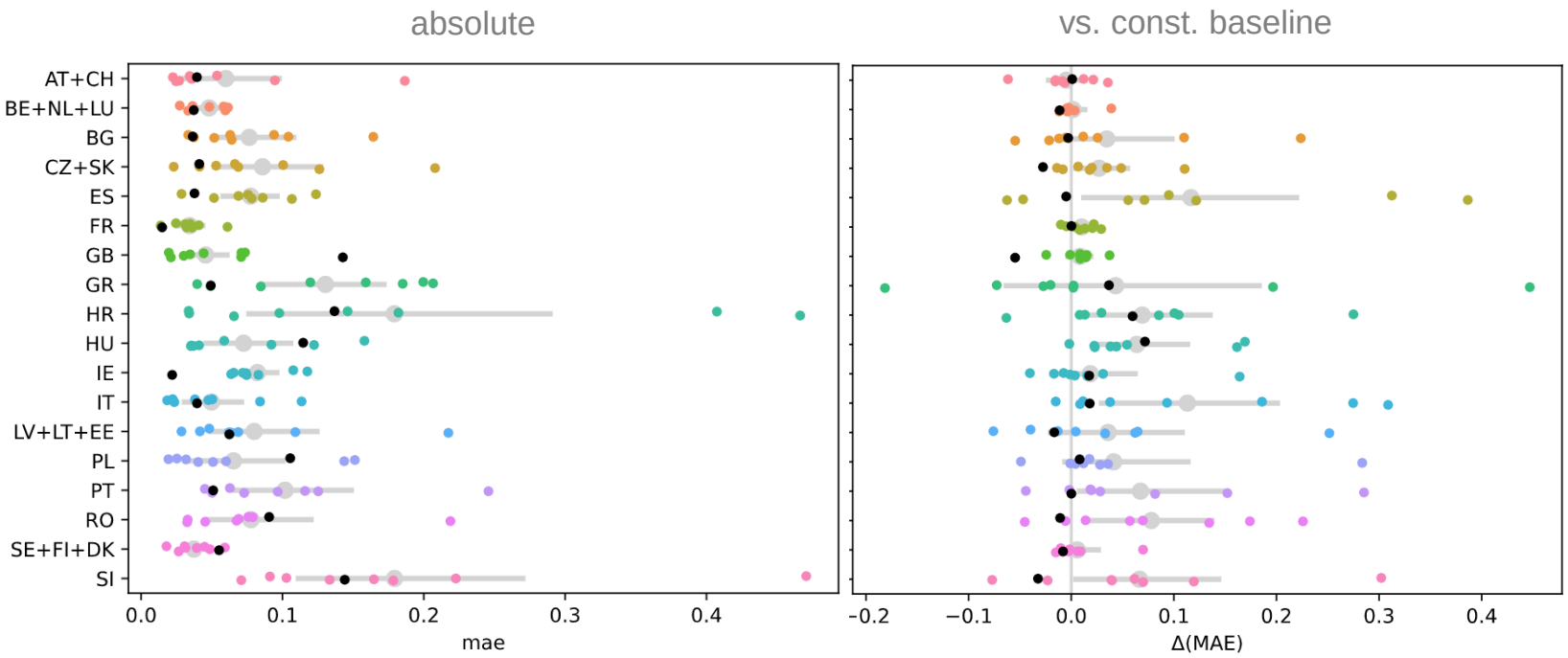
Model: Random Forest
Combine test point predictions from CV; grey area: holdout set

Choice of keywords: only KW group 19 useful to predict registrations



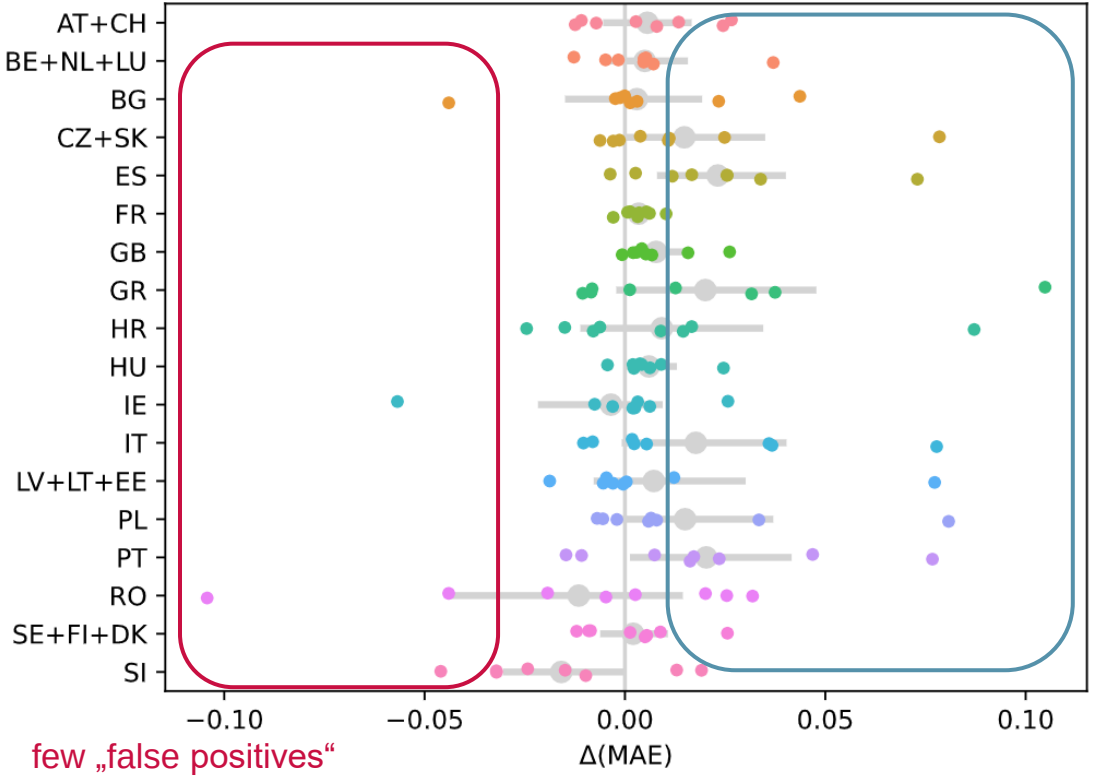
- Random forest, only lagged values of specified Google Trends index as features
- KW group 19: „jobs germany“/“work germany“ (cf. Wanner, 2020, https://doi.org/10.1007/s11135-020-01047-w)
- Adding *k* best extra keyword groups -> only minimal improvement

For which countries does the prediction work well?



points: CV test folds (one per year) **black points**: holdout set grey circles and lines: mean, 95% C.L.

With vs. without Google Trends: slightly better predictions in a few cases



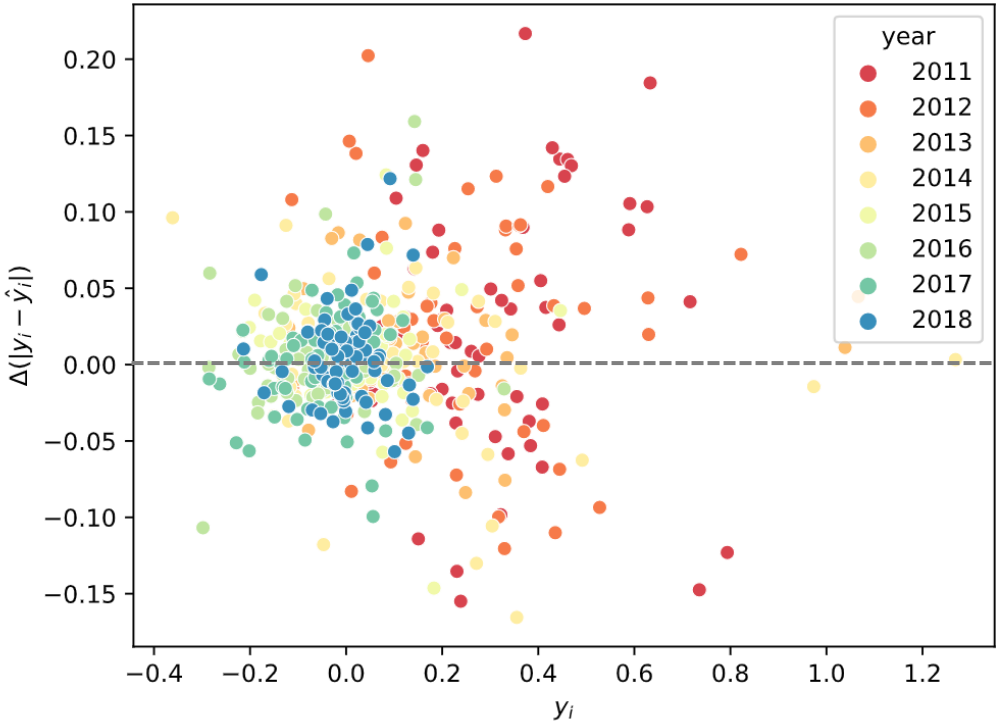
few „false positives“

up to ~ 5 – 10 % error decrease

points: CV test folds (one per year)
grey circles and lines: mean, 95% C.L.

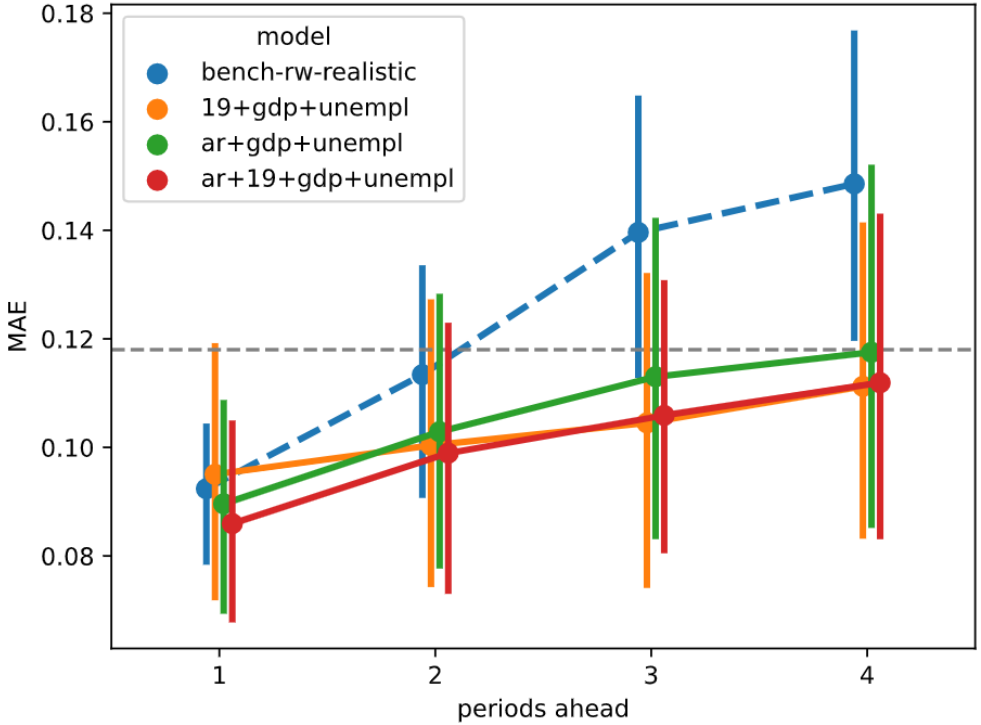
Google Trends most beneficial for predicting growth in early 2010s

Forecast error
no Trends vs. Trends



true value

N-step ahead forecast: performance largely stable, slight advantage over autoregression the long run



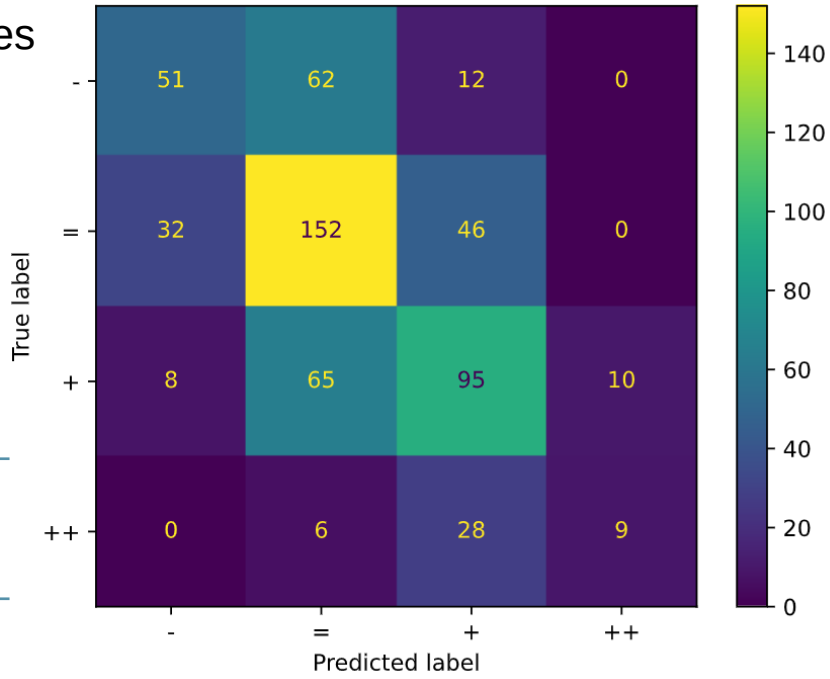
Benchmark: previous(2)

Autoregression + Eurostat
Benchmark: const. 0
Trends + Eurostat
All combined

Random Forest, 95% C.L. errorbars

Could a classification approach be an alternative application?

- Discretize next quarter forecast into 4 (3) classes (-, =, +, (++))
- Model: Random Forest Classifier
- Tune boundaries while ensuring sufficient class members
- Result: insufficient distinction of neighbouring labels (esp. - and ++), but few egregious errors

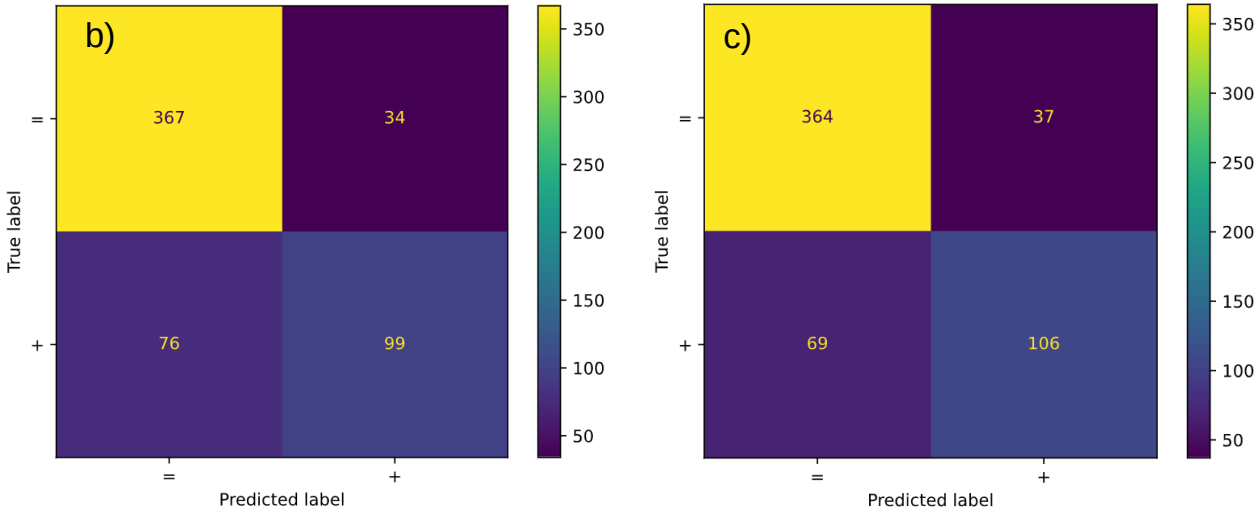


Classes	Micro Acc.	F1 macro	Precision macro	Recall macro	Earth Movers Distance
4	0.53	0.41	0.42	0.45	0.29
3	0.59	0.46	0.48	0.50	0.24

Mean CV scores

Binary classification („excess forecast“) as minimalistic solution

- Pure Trends model again nearly on par with autoregression
- Minimal improvement by combination
- N.b: positive classes concentrated on 2011 - 2016



Features	AUC	Accuracy	F1	Precision	Recall
a) Autoregressive + Eurostat	0.80	0.73	0.58	0.66	0.56
b) Trends + Eurostat	0.78	0.76	0.53	0.65	0.48
c) All combined	**0.71**	**0.77**	**0.62**	**0.70**	**0.58**

Mean CV scores 2011 - 2016

Theoretical Insights

- Strong and stable seasonality of EU migration towards Germany, forecasts need to focus on non-cyclic component
- Google Trends in principle valuable for short to medium term migration forecast
- Data quality is still a huge issue
- Right now, Trends can be used to partly improve forecast accuracy and longterm stability

Practical Learnings

- Google is a blackbox
- The forecasters‘ toolbox differs from the explainers‘ toolbox
- A weather forecast is different from a climate forecast

Discussion points

- Are there further models (both classical and bleeding-edge) that can be compared with the Google Trends approach?
- What is missing from the Google Trends approach in order to achieve continuous and verifiable better forecast results?
- What can the Google Trends approach in its current state be used for?

Thank you!

Besuchen Sie uns auch auf

www.bertelsmann-stiftung.de

Backup

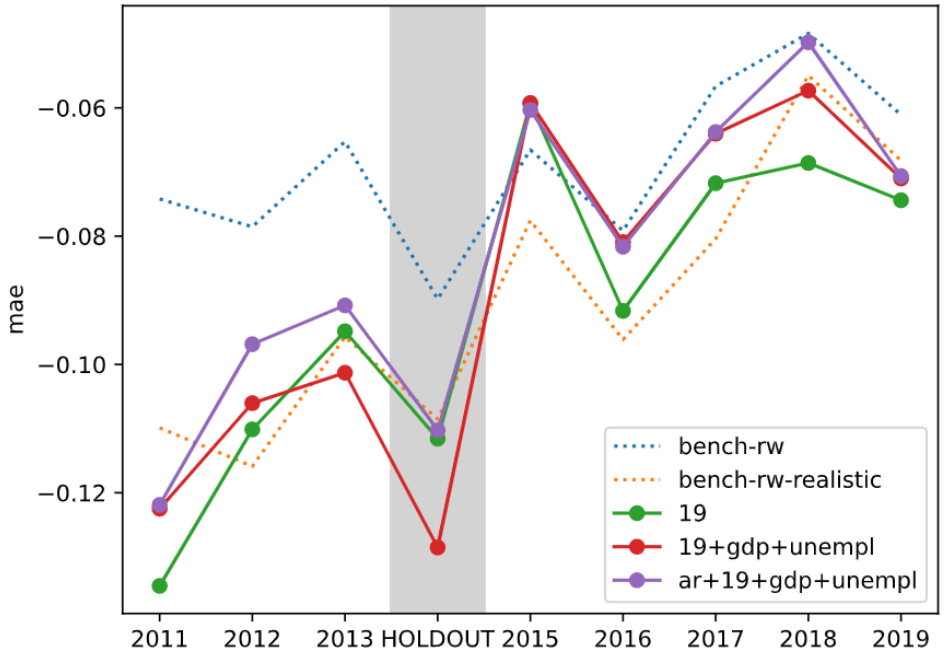
BertelsmannStiftung

Predicting next **month**: results similar to predicting next quarter

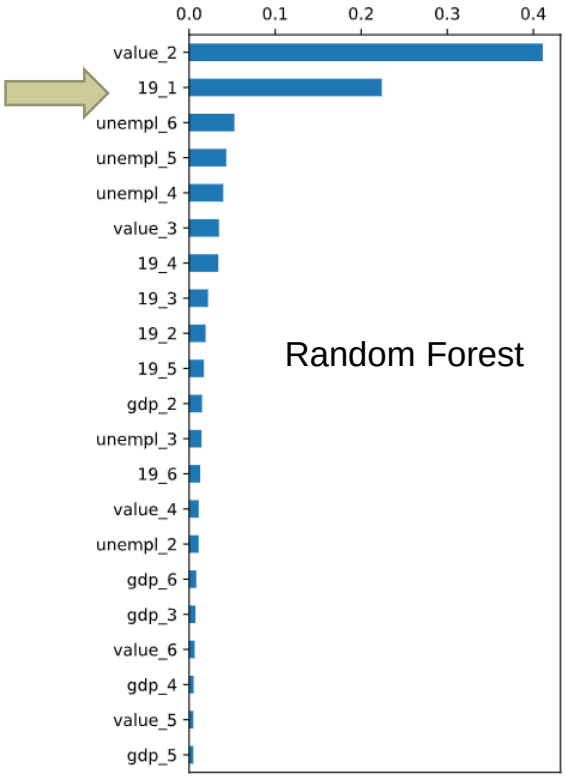
Model	Lagged Features *	R^2_{OOS} **	RMSE **	MAE **
Benchmark: const. 0	-	0.0	0.185	0.137
Benchmark: previous(1)	-	**0.15**	**0.143**	**0.107**
Benchmark: previous(5)	-	-0.09	0.172	0.127
OLS	self	0.30	0.145	0.104
OLS	Google Trends (Keyword 19)	0.14	0.158	0.110
OLS	combined	**0.36**	0.138	**0.097**
Random Forest	self	0.26	0.149	0.107
Random Forest	Google Trends (Keyword 19)	0.18	0.152	0.108
Random Forest	combined	0.34	**0.137**	0.098

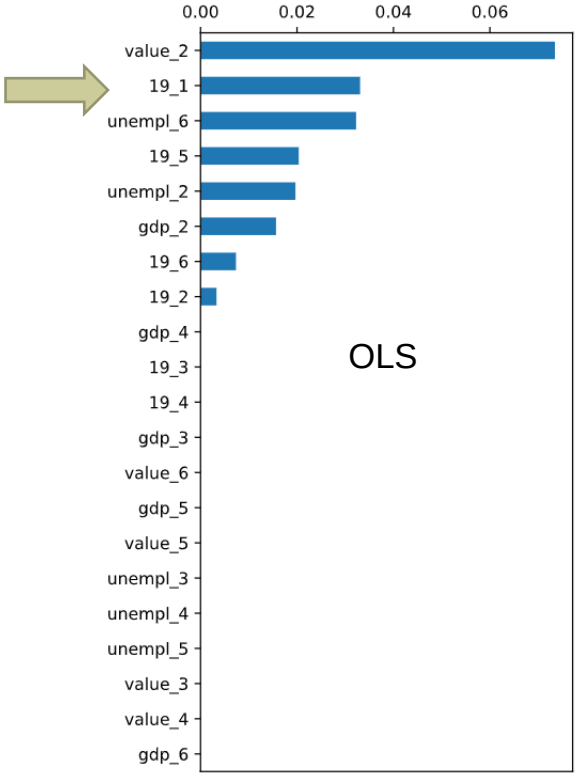
*) up to 18 periods, self: min. delay of 2 periods, GDP/unemployment: min. delay of 1 period;
**) mean performance on cross-validation test folds

Performance by year: Google beneficial during Euro crisis years and during mid-2010s, but improvement vs benchmark not huge

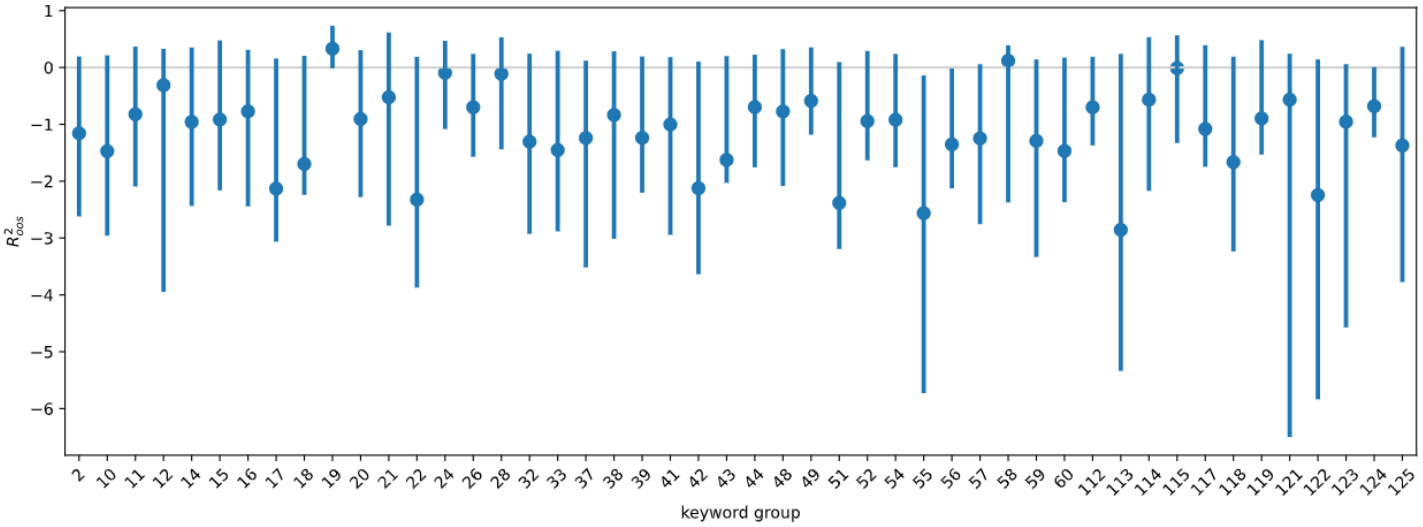


First lag of Google Trends index second most important feature





What about splitting the model into macroregions?

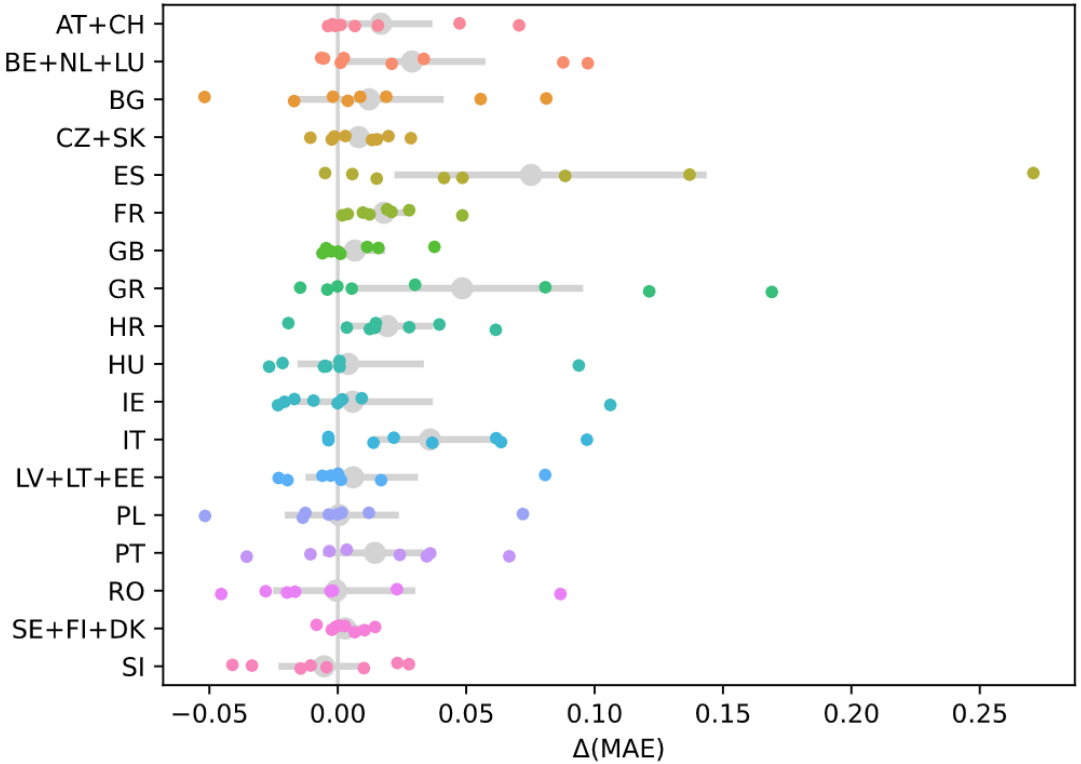


- E.g. South Europe (ES, IT, PT, GR): more keywords are possible
- Best keyword combination (select *k* best): 19, 115, 28, 12 (“jobs germany”, “unemployment”, “vacancies germany”, “minimum wage germany”)

But compared individually, no clear benefit from regional model

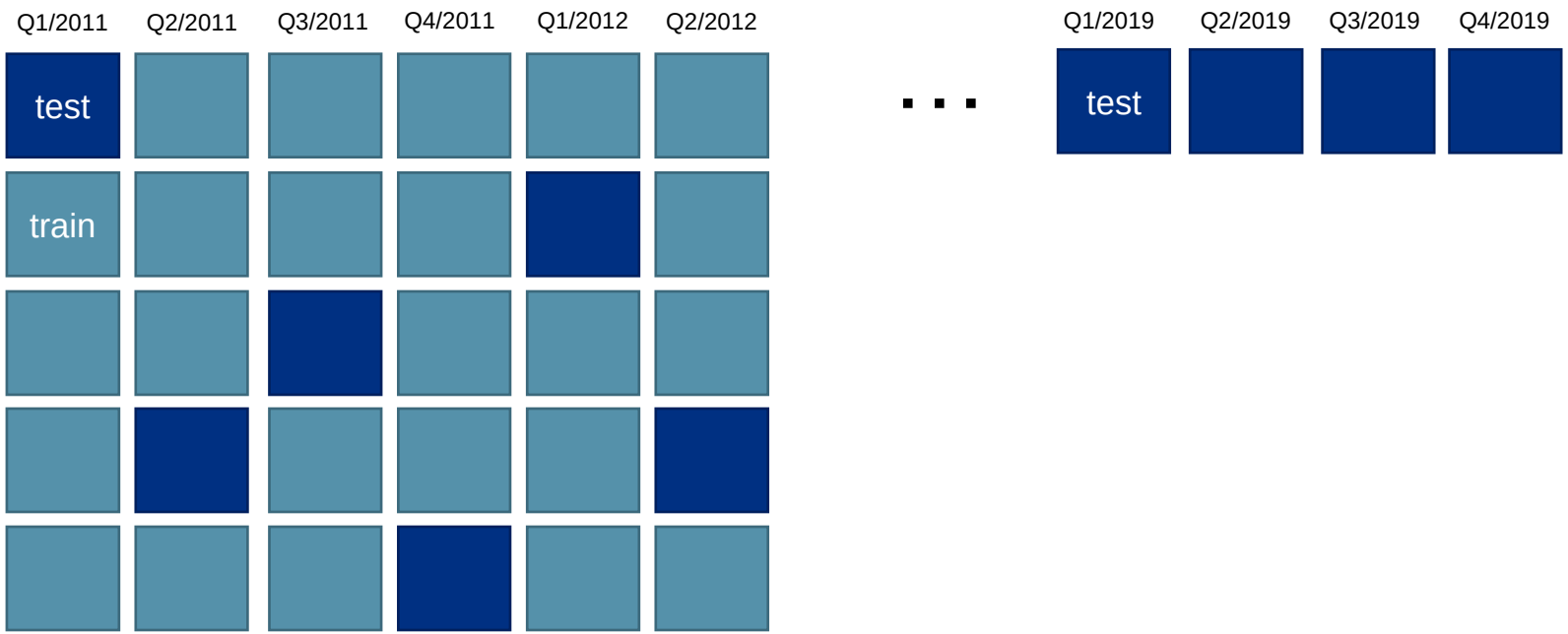
	Regional model		Global model	
country	**MAE**	**RMSE**	**MAE**	**RMSE**
GR	**0.101**	0.117	0.102	**0.116**
ES	0.084	0.092	**0.079**	**0.088**
PT	**0.071**	**0.084**	0.087	0.097
IT	0.073	0.079	**0.066**	**0.074**

Eurostat with vs without Trends: mostly clear improvement



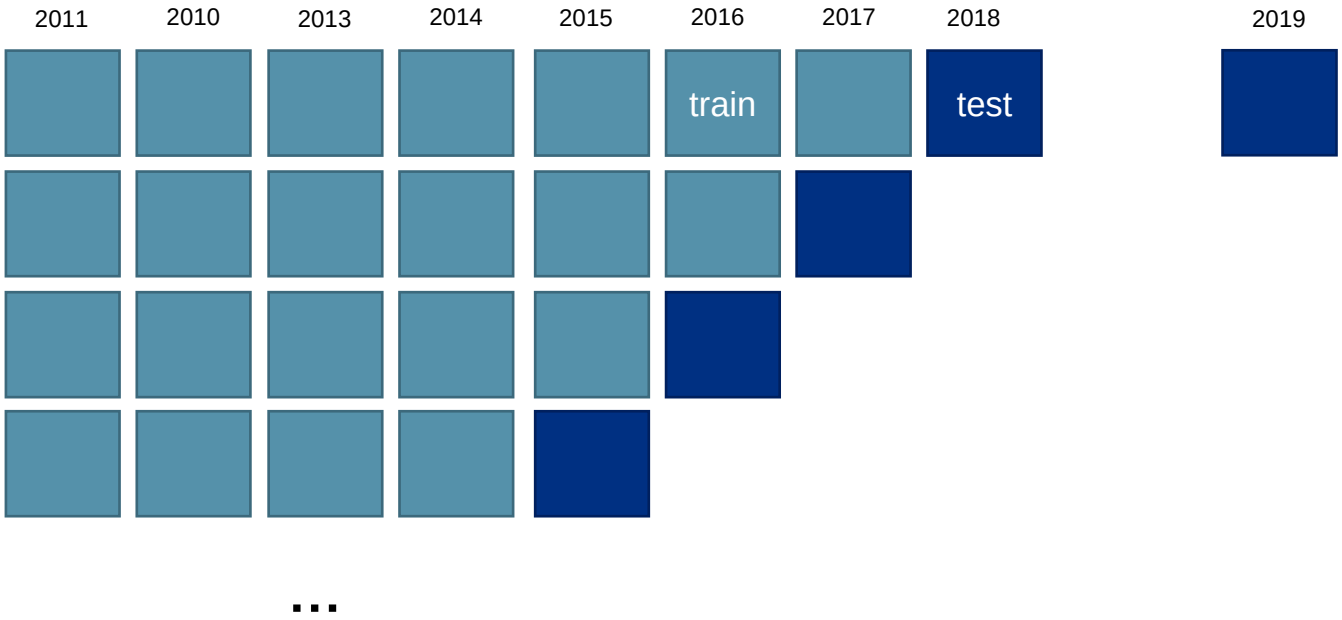
points: CV test folds (one per year)
grey circles and lines: mean, 95% C.L.

k-fold crossvalidation + out-of-sample eval. for consistency checks



Allowed for most time-series regression models
(Bergmeir et al., 2018, https://robjhyndman.com/publications/cv-time-series/)

Time series cross validation + out-of-sample eval. for consistency checks



Can also use k-fold CV but only of residuals are uncorrelated
(Bergmeir et al., 2018, https://robjhyndman.com/publications/cv-time-series/)