ESSnet Big Data

Specific Grant Agreement No 1 (SGA-1)

https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata
http://www.cros-portal.eu/.........

Framework Partnership Agreement Number **11104.2015.006-2015.720**

Specific Grant Agreement Number **11104.2015.007-2016.085**

Work Package 6

Early Estimates

Deliverable 6.3

Recommendations about methodology for processing the data for purposes of Consumer Confidence Index and NowCasts of Turnover Indices

Version 2016-16-06

Prepared by:

Henri Luomaranta (Statistics Finland)
Piet Daas (CBS, Netherlands)
Anna Nowitzka (GUS, Poland)
Boro Nikic (SURS, Slovenia)

ESSnet co-ordinator:

Peter Struijs (CBS, Netherlands)
p.struijs@cbs.nl
telephone : +31 45 570 7441
mobile phone : +31 6 5248 7775

Table of contents

1. INTRODUCTION

The aim of the deliverable "Recommendations about methodology for processing the data for purposes of Consumer Confidence Index and NowCasts of Turnover Indices" is to offer a range of methodological recommendation related to methodology being investigates for purposes of nowcast of early economic indicators with possibly inclusion of combined data sources.

During the ESSNet project in the SGA-1 period some nowcasting methods for the purpose of estimating Turnover Indices were discovered and tested. The original purpose was to explore the possibilities of estimating the consumer confidence index and (or) indices of turnover. Due to the early findings about the inaccessibility of social media data, which is crucial for assessing the consumer confidence index, the WP6 team focused on Turnover Indices.

This report contains findings which have been discovered estimating Turnover Indices mostly at Statistics Finland and Statistical office of the Republic of Slovenia.

2. NOWCASTING

Nowcasting is a very early estimate produced for an economic variable of interest over the most recent reference period calculated on the basis of incomplete data using a statistical or econometric model different from the one used for regular estimates. Soft data should not play a predominant role in nowcasting models. Nowcasts may be produced during the very same reference period for which the data is produced.

Conducting one of the pilots during SGA1 ESSnet project several nowcasting methods were under investigation. Among them the most promising in sense of practical implementation was Principal components analysis method. The central idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables.

3. PRINCIPAL COMPONENT ANALYSIS (SLOVENIAN EXPERIENCES)

The main idea is to (practically) test at least one of the nowcast methods for purposes of estimating early economic indicators. As it was mentioned the PCA (principal component analysis) model has been tested at Statistical Office of the Republic of Slovenia for these purposes.

PCAModel : [i]consistes of two stages:

1. Principal component analysis (PCA)

- dimensionality reduction
- time series of (enterprise) data → standardize → choose the first few principal components
- various conditions for choosing principal components:
    - ❖ The chosen principal components explain at least 70% (75%, 80%, 85%, 90%) of variability of enterprise data.
    - ❖ Time series in the linear regression model are at least 7 (8, 10, 15, 20) times longer than the number of the chosen principal components.
    - ❖ The last chosen principal component explains at least 5% of variability of enterprise data.

2. Linear regression

- Y (dependent variable): time series of interest, e.g. turnover index

- $x_1, x_2, \cdots, x_k$, (predictors): e.g. the chosen principal components

SURS With the help of statistics Finland prepared application (together with instructions how to use it) which allow:

- inputting different kind of data
- testing various conditions for choosing principle components
- producing quality indicators which compare results of different nowcasting methods
- producing quality indicators which compare results with disseminated official statistics

For testing the following possibilities for models were considered:

- Time series of interest: GDP in constant prices (chain-linked volumes, reference year 2010) from 2008Q1 to 2015Q4.
- 8 different testing spans:
  - the first period is always 2008Q1;
  - the last period is 2014Q1 or 2014Q2 or … or 2015Q4.
- 3 different sets of enterprise data:
  - D1: turnover in industry[1];
  - D2: turnover in retail trade[2];
  - D3: turnover in industry and retail trade (i.e. D1 and D2 together).
- 5 different conditions for choosing enterprise data:

Table 1: Conditions for choosing enterprise data

Condition	Meaning
s10	choose only raw enterprise data that are available sooner than 10 days (i.e. until the end of the 9th day) after the end of the last period
s20	choose only raw enterprise data that are available sooner than 20 days (i.e. until the end of the 19th day) after the end of the last period
s30	choose only raw enterprise data that are available sooner than 30 days (i.e. until the end of the 29th day) after the end of the last period
s46	choose only raw enterprise data that are available sooner than 46 days (i.e. until the end of the 45th day) after the end of the last period
u	choose edited enterprise data

- 11 different conditions for choosing principal components:

Table 2: Conditions for choosing principal components

Condition	Meaning
zadnja5	take every p. c., whose eigenvalue's share among all eigenvalues is greater or equal to 5%

[1] The data set is a good approximation of the real state.

[2] The data set is a good approximation of the real state.

po7	take only as many p. c. to have at least 7 cases (time periods) per independent variable later in the linear regression
po8	take only as many p. c. to have at least 8 cases (time periods) per independent variable later in the linear regression
po10	take only as many p. c. to have at least 10 cases (time periods) per independent variable later in the linear regression
po15	take only as many p. c. to have at least 15 cases (time periods) per independent variable later in the linear regression
po20	take only as many p. c. to have at least 20 cases (time periods) per independent variable later in the linear regression
70	take enough p. c. to explain 70% (or a bit more) of the variability of the enterprise data
75	take enough p. c. to explain 75% (or a bit more) of the variability of the enterprise data
80	take enough p. c. to explain 80% (or a bit more) of the variability of the enterprise data
85	take enough p. c. to explain 85% (or a bit more) of the variability of the enterprise data
90	take enough p. c. to explain 90% (or a bit more) of the variability of the enterprise data

- Seasonality[3] can be added as an additional predictor or not (yes or no).
- Sentiment indicator[4] can be added as an additional predictor or not (yes or no).

Altogether 5280[5] models are made.

Comments on data preparation:

- Enterprise data are prepared using SAS.
- The data sets used are a good approximation of the real state. It is impossible for us to get a true state for a certain data set at a certain time in the past, but we can estimate the state well.
- Since we started using e-questionnaires[6] (2013M04 in industry, 2014M01 in retail trade), we have the data for some enterprises available only a few days after the end of the reference period. So we are able to get early estimates based on these data.

Comments on models:

[3] Seasonal component for the time series of interest is calculated using function stl in package stats.

[4] The computation of Sentiment indicator is different from the computation of Economic sentiment indicator by European Commission, but they are highly correlated. The Sentiment indicator used for testing is lagged (e.g. values for 2008M01, 2008M02, 2008M03 are used to predict the time series of interest for period 2008Q2).

[5] $5280 = 8 \cdot 3 \cdot 5 \cdot 11 \cdot 2 \cdot 2$

[6] Electronic questionnaires.

- Models are prepared using R. R packages used for testing are: stats, graphics, missMDA, Formula, Hmisc, foreign, ggplot2, gvlma, car, lmtest, haven, dplyr, reshape.
- The reason condition s46 for choosing enterprise data is chosen, is that VAT data used for turnover in retail trade is available at t+45.
- One of the possible conditions for choosing principal components could be to take every principal component, whose eigenvalue is greater or equal to 1. But in this way, often too many principal components are chosen, which is not appropriate for linear regression.
- Linear regression could be further optimized, but our models produced some errors, so we excluded this from testing.

Results of testing

Several diagnostics for each of 5280 models are computed. Also, several diagnostics are computed for each group of 8 models with different testing spans but the same conditions (e.g. D1, s20, po7, seasonality = yes, sentiment indicator = yes).

To evaluate the predictions (nowcasts) of the models, the diagnostic chosen for comparison of the models is the mean squared error of the nowcasts[7] (MSE) computed for each group of 8 models mentioned above.

As benchmark, the model with specification selected automatically by JDemetra+ version 2.1.0 is used.[8] The forecasts are obtained and the mean squared error of the forecasts (MSE) is computed.

To evaluate how much the predictions (nowcasts) deviate from the true values, the mean absolute error of the nowcasts[9] (MAE) is computed for each group of 8 models mentioned above.

Table 3: Number of enterprises

Question	**D1[10]**	**D2[11]**	**D3[12]**
How many enterprises have data on the last period (on average)?[13]			
→ s10	164	168	332
→ s20	730	235	966
→ s30	1303	236	1539
→ s46	1325	2380	3705
→ u	1600	2825	4425

[7] #diag_napaka_meansq
[8] #a1
[9] #diag_napaka_meanabs
[10] #analiza3_v01_(primer4).R
[11] #analiza3_v01_(primer5).R
[12] #analiza3_v01_(primer6).R
[13] #st_podj_konec

How many enterprises have data on the last period and on the rest of testing span (on average)?[14] (share)			
→ s10	17 (10%)	26 (16%)	43 (13%)
→ s20	74 (10%)	36 (15%)	110 (11%)
→ s30	88 (7%)	36 (15%)	124 (8%)
→ s46	88 (7%)	1159 (49%)	1247 (34%)
→ u	1017 (64%)	1615 (57%)	2632 (59%)

Table 3 shows that there are only a few hundred enterprises that reported sooner than 10 days after the end of the reference period. But only a little over 10% of them can be used for the model, because such enterprises have to have the data for every period of the testing span. In time, more enterprises report but still only a small percentage of them can be used for the model; the exception is s46, because at t+45 turnover data in retail trade are obtained for many enterprises from VAT. In case of edited data we are able to include around 60% of the enterprises to the model.

Table 4: Influence of additional predictors and comparison to benchmark model (MSE is used)

Question	**D1[15]**	**D2[16]**	**D3[17]**
How often is a model with seasonality as an additional predictor better than the same model without it?	97.3%	89.1%	92.7%
How often is a model with sentiment indicator as an additional predictor better than the same model without it?	94.5%	94.5%	93.6%
How often is a model better than the benchmark model?	32.7%	29.1%	32.3%
How often is a model with seasonality and sentiment indicator as additional predictors better than the benchmark model?	96.4%	83.6%	85.5%

Table 4 shows that seasonality as an additional predictor improves the model in most cases. Similarly, sentiment indicator as an additional predictor improves the model in most cases. Models with seasonality and sentiment indicator as additional predictors are better than the benchmark model in more than 80% or even more than 90% of the cases.

[14] #st_podj_podatki
[15] #analiza1_v02_(primer4).R
[16] #analiza1_v02_(primer5).R
[17] #analiza1_v02_(primer6).R

Table 5: Comparison among datasets (MSE is used)

Question	
How often is a model based on D1 better than the same model based on D2?[18]	73.6%
How often is a model based on D1 better than the same model based on D3?[19]	65.0%
How often is a model based on D2 better than the same model based on D3?[20]	29.1%
How often is a model with seasonality and sentiment indicator as additional predictors based on D1 better than the same model based on D2?[21]	45.5%
How often is a model with seasonality and sentiment indicator as additional predictors based on D1 better than the same model based on D3?[22]	49.1%
How often is a model with seasonality and sentiment indicator as additional predictors based on D2 better than the same model based on D3?[23]	52.7%

Table 5 shows that, considering all models, data set D1 gives better models than D3 which gives better models than D2. But considering only models with seasonality and sentiment indicator as additional predictors, all three data sets give similar results.

Table 6: (MSE is used)

D1[24]	All models	Only 'good' models[25]
How often is a model with condition s20 better than the same model with condition s10?	45.5%	9.1%
How often is a model with condition s30 better than the same model with condition s20?	68.2%	54.5%
How often is a model with condition s46 better than the same model with condition s30?	they are the same	they are the same
How often is a model with condition u better than the same model with condition s10?	52.3%	27.3%
How often is a model with condition u better than the same model with condition s20?	54.5%	54.5%

18 #analiza2_v01_(primer45).R
19 #analiza2_v01_(primer46).R
20 #analiza2_v01_(primer56).R
21 #analiza2_v01_(primer45).R
22 #analiza2_v01_(primer46).R
23 #analiza2_v01_(primer56).R
24 #analiza4_v01_(primer4).R
25 Only models with seasonality and sentiment indicator as additional predictors.

How often is a model with condition u better than the same model with condition s30?	50.0%	45.5%
How often is a model with condition u better than the same model with condition s46?	50.0%	45.5%

Table 7: (MSE is used)

D2[26]	All models	Only 'good' models[27]
How often is a model with condition s20 better than the same model with condition s10?	72.7%	54.5%
How often is a model with condition s30 better than the same model with condition s20?	they are the same	they are the same
How often is a model with condition s46 better than the same model with condition s30?	54.5%	18.2%
How often is a model with condition u better than the same model with condition s10?	61.4%	27.3%
How often is a model with condition u better than the same model with condition s20?	52.3%	18.2%
How often is a model with condition u better than the same model with condition s30?	52.3%	18.2%
How often is a model with condition u better than the same model with condition s46?	40.9%	27.3%

Table 8: (MSE is used)

D3[28]	All models	Only 'good' models[29]
How often is a model with condition s20 better than the same model with condition s10?	75.0%	63.6%
How often is a model with condition s30 better than the same model with condition s20?	56.8%	45.5%
How often is a model with condition s46 better than the same model with condition s30?	56.8%	45.5%

[26] #analiza4_v01_(primer5).R
[27] Only models with seasonality and sentiment indicator as additional predictors.
[28] #analiza4_v01_(primer6).R
[29] Only models with seasonality and sentiment indicator as additional predictors.

How often is a model with condition u better than the same model with condition s10?	79.5%	54.5%
How often is a model with condition u better than the same model with condition s20?	56.8%	45.5%
How often is a model with condition u better than the same model with condition s30?	65.9%	54.5%
How often is a model with condition u better than the same model with condition s46?	77.3%	63.6%

Table 9: Sum of ranks for all models for different conditions for choosing principal components (MSE is used)

Sum of ranks for all models	**D1**[30]	**D2**[31]	**D3**[32]
zadnja5	130	97	106
po7	126	120	84
po8	108	114	109
po10	143	127	162
po15	189	186	191
po20	163	181	180
70	87	84	87
75	80	84	73
80	71	82	52
85	91	121	103
90	132	124	173

We are able to rank models with different conditions for choosing principal components according to the value of mean squared error of the nowcasts. Table 9 shows that, considering all models, condition 80 for choosing principal components has the lowest sum of ranks, so from this perspective it is the best. Conditions 75 and 70 also seem very good. Conditions po15 and po20 have the highest sum of ranks, so from this perspective they are the worst.

[30] #analiza5_v01_(primer4).R
[31] #analiza5_v01_(primer5).R
[32] #analiza5_v01_(primer6).R

Table 10: Average of MAE

Average of MAE[33]	All models	Only: season. = no sent. ind. = no	Only: season. = yes sent. ind. = no	Only: season. = no sent. ind. = yes	Only: season. = yes sent. ind. = yes
D1[34]					
→ s10	169.0	242.1	204.3	153.2	76.2
→ s20	161.8	226.6	153.4	177.3	90.0
→ s30	155.9	209.7	152.1	169.4	92.5
→ s46	155.9	209.7	152.1	169.4	92.5
→ u	150.9	203.9	156.3	156.1	87.1
D2[35]					
→ s10	241.4	389.6	225.7	280.9	69.3
→ s20	218.1	352.2	208.1	237.3	74.9
→ s30	218.1	352.2	208.1	237.3	74.9
→ s46	176.3	243.8	229.7	137.3	94.3
→ u	185.5	250.1	235.6	149.7	106.6
D3[36]					
→ s10	208.0	291.3	224.6	221.6	94.6
→ s20	174.9	258.0	179.7	176.6	85.4
→ s30	177.7	259.3	181.5	180.0	89.8
→ s46	177.9	245.1	227.0	141.1	98.7
→ u	167.2	219.1	200.0	149.5	100.9

Table 10 shows that, considering all models, average of MAE is smaller if we have more data (i.e. more days after the end of the reference period). In case of D1 and D3, edited data have the lowest average of MAE. But if we focus on models with/without seasonality and/or with/without sentiment

[33] The model with the lowest value of MAE. It has also the lowest value of MARE.
[34] #analiza7_v01_(primer4).R
[35] #analiza7_v01_(primer5).R
[36] #analiza7_v01_(primer6).R

indicator as additional predictors, the situation is different. We can see that models with seasonality and sentiment indicator as additional predictors have the lowest average of MAE.

Table 11: The best model according to MAE: its MAE, MARE, MaxAE, MaxARE

The best model according to MAE[37]	MAE[38]	MARE[39]	MaxAE[40]	MaxARE[41]
D1[42]				
→ s10	61.9	0.69%	146.8	1.58%
→ s20	67.9	0.74%	121.5	1.31%
→ s30	63.7	0.70%	121.7	1.31%
→ s46	63.7	0.70%	121.7	1.31%
→ u	54.6	0.61%	114.2	1.34%
D2[43]				
→ s10	39.2	0.44%	94.6	1.11%
→ s20	47.4	0.53%	119.4	1.27%
→ s30	47.4	0.53%	119.4	1.27%
→ s46	65.9	0.72%	212.3	2.29%
→ u	73.3	0.80%	164.0	1.77%
D3[44]				
→ s10	80.8	0.89%	171.3	1.85%
→ s20	68.8	0.75%	108.5	1.17%
→ s30	69.4	0.76%	165.9	1.79%
→ s46	65.9	0.73%	181.5	1.96%
→ u	67.6	0.73%	188.6	2.03%

[37] The model with the lowest value of MAE. It has also the lowest value of MARE.
[38] MAE = mean absolute error of the nowcasts
[39] MARE = mean of absolute values of relative errors of the nowcasts
[40] MaxAE = maximum absolute error of the nowcasts
[41] MaxARE = maximum of absolute values of relative errors of the nowcasts
[42] #analiza6_v01_(primer4).R
[43] #analiza6_v01_(primer5).R
[44] #analiza6_v01_(primer6).R

Table 11 shows us that the best models according to MAE (and also MARE) have mean absolute relative error (MARE) below 1%. But maximum absolute relative error (MaxARE) could be over 2%. We are able to find good models even 10 days after the end of the reference period.

We can conclude that we get some very good models that can be used for nowcasting. But although we use a model with, for example, very low MAE and MSE, this doesn't ensure us a small nowcasting error, but rather gives us hope for it.

We think that it would be better to have more testing spans so that more values would be used for calculation of the averages and therefore the diagnostics would be more realistic.

Ideas for the future

There are still some possibilities to improve the nowcasting model, for example:

- We could try to have more testing spans so more values would be used for calculation of averages etc.
- 10 days after the end of the reference period we already have raw turnover data from few hundred enterprises. But there are only few enterprises that have the data for every period of our testing span. We could try to impute or estimate the missing data from some enterprises to get more data for our model.
- We could try to prepare more sets of enterprise data, for example for services, and test how this set influences the results.
- We could add some other predictors to linear regression, for example a predictor derived from traffic sensors' data, a predictor derived from job vacancies.
- We are in progress of making an agreement to receive VAT data sooner than at t+45.
- We could further investigate the possibilities for optimization of linear regression.

4. NOWCASTING TURNOVER INDEXES (FINNISH EXPERIANCE)

The main reasons of work done by Statistics Finland were:

- Pressing issue was the long lag of publication and requirements from FRIBS and users (i.e. national accounts)
- StatFi wanted a practical solution, so another aim was simplicity and tractability in terms of data source and method, if possible
- To propos methods that can be useful in Big Data, and that are commonly used in the Nowcasting of macroeconomic variables
- Using continuously accumulating firm level data (hard data)
- Estimating the common components underlying this data with factor analysis
- The common components would be predictors in nowcasting equations

- Looking at the machine learning literature, other options were available (such as LASSO, RIDGE, Elastic Net regressions) that could deal with high dimensional econometric problems

The firm level data was used for those purposes. One of the issues which had to be solved was multidimensionality of data. Widely used is the factor analysis which estimates the common and idiosyncratic variance underlying the data. But shocks to large companies can have a sizeable impact in the Finnish economy (economic activity is concentrated on a few multinationals e.g. Nokia). That is why so-called shrinkage models were explored in order to better capture some of the firm specific variations. They include all the firm growth rates in the estimation, and deal with the curse of overfitting by shrinking parameter values towards 0. These models outperform factor models (in general).

Table 12: Results 1st strategy MAE – results ar better than with ARIMA benchmark

Table 1: MAEs for the nowcasts of year-on-year growth rates of the turnovers indexes, obtained using firm-level data. Results are in percentage points and bolded numbers indicate the lowest MAE for a given estimation period. The bottom row is the MAE of the $t|45$ estimate by the statistics office.

	t—5	t—10	t—20	t—26
		Manufacturing		
Elastic Nets	2.097	2.324	2.017	1.999
Lasso	2.268	2.369	2.177	2.122
Ridge	**1.833**	**1.861**	**1.712**	**1.675**
Factors	2.739	2.582	2.166	2.036
t—45 (current method)	**1.084**			
		Services		
Elastic Nets	1.471	1.248	**0.979**	**0.984**
Lasso	1.508	1.251	0.980	0.999
Ridge	**1.269**	**1.212**	1.071	1.076
Factors	1.707	1.348	1.183	1.166
t—45 (current method)	**1.730**			
		Trade		
Elastic Nets	1.978	1.697	**1.165**	**1.162**
Lasso	2.045	1.758	1.172	1.175
Ridge	**1.762**	**1.650**	1.292	1.297
Factors	2.272	1.726	1.441	1.443
t—45 (current method)	**0.895**			
		Construction		
Elastic Nets	3.302	3.620	2.851	3.054
Lasso	3.358	3.641	2.800	3.098
Ridge	**2.982**	3.308	3.029	2.988
Factors	3.466	**3.215**	**2.487**	**2.359**
t—45 (current method)	**2.555**			

Table 13: Results 2nd strategy MAE – and we obtain accurate results

Table 1: MAEs for the nowcasts of year-on-year growth rates of the turnovers indexes at $t|28$, obtained by predicting the VAT component with the accumulated firm level data, using 1 and 2 factors as inputs in an ARIMA model. Results are in percentage points.

	Manufacturing	Services	Trade	Construction
MAE of estimate (1 factor)	0.592	0.391	0.535	1.100
MAE of estimate (2 factors)	0.423	0.381	0.543	2.295
MAE of estimate no factors	0.844	0.917	0.751	1.914
MAE of $t|45$ (current method)	1.084	1.730	0.895	2.555

Statistics Finland has also tested nowcasting the second month of the quarter and forecasting the third, allowing to compute real time estimates of quarterly GDP. Methods were similar, using sales inquiry for firm level data. Turnover indexes are widely followed in their own right, but are also used as source material for producing the Trend Indicator of Output (TIO, i.e. the Finnish monthly economic activity indicator).

More detailed report about results obtained by Statistics Finland in the separated document

D6_3 Stat Finland turnovers_nowcast_project_bis.pdf

5. CONCLUSIONS

The main goal of SGA-1 period was to achieve at least one "quick win" in testing so called big data methods in calculating early estimates of Consumer Confidence Index and (or) NowCasts of Turnover Indices. WP 6 team managed to deeply analyses several methods and use them in order to calculate early estimates of turnover indicators and also prepare quality measures of calculated estimates in order to compare them with other models (among nowcasting models, same models with different parameters, ARIMA models…). These results are encouraging even if just data from traditional sources were used, however the plan for SGA-2 is to combine this data with some of big data sources and test various methodological issues on selected early economic indicators. The additional plan for SGA-2 is also to further investigate of rapid estimators [45] which could be useful for business case described in the deliverable D1.

[45] https://unstats.un.org/unsd/nationalaccount/consultationDocs/Handbook_RE.pdf