CEDEFOP

European Centre
for the Development
of Vocational Training

40 years

ORGANISING

EU-SBAS INFORMATION DAY

A SHORT-TERM BASED ANTICIPATORY SYSTEM FOR THE EUROPEAN LABOUR MARKET TRENDS, SKILLS DEVELOPMENTS AND VET POLICY

SBAS
A SECTORAL BASED ANTICIPATORY SYSTEM FOR THE EUROPEAN LABOUR MARKET

AN OVERVIEW OF THE METHODOLOGY

PROF. MASSIMILIANO MARCELLINO

13/02/2015

Table of Contents

- **Data**

- **Models**

- **Forecasting Exercise**

- **Overview of Results**

- **Conclusions**

DATA: DEPENDENT VARIABLES

- **DEPENDENT VARIABLES IN THREE PHASES:**

1. SERIES FOR EACH EU COUNTRY DISAGGREGATED BY OCCUPATION, BASED ON ISCO, FURTHER DIVIDED INTO MALES AND FEMALES AND INTO FOUR AGE GROUPS: 15-24, 25-49, 50-64 AND 65+ (ABOUT 4000 SERIES)

2. DISTRIBUTE THE OCCUPATION SERIES INTO SECTORS, BASED ON NACE, USING A NACE/ISCO IMPUTATION MATRIX PROVIDED BY CEDEFOP/EUROSTAT. FORECASTING SECTORAL EMPLOYMENT DEVELOPMENTS FOR EACH EUROPEAN COUNTRY, DISAGGREGATED BY SEX AND AGE GROUP (ABOUT 3500 SERIES)

3. FOLLOWING CEDEFOP, WE DISTRIBUTE THE OCCUPATIONS INTO LOW, MEDIUM AND HIGH SKILLS. THIS PROVIDES SHORT TERM SKILLS FORECASTS FOR EACH COUNTRY, AGAIN DISAGGREGATED BY SEX AND AGE (ABOUT 470 SERIES)

Data: Independent Variables

■ INDEPENDENT VARIABLES:

✓ About 50 variables for each of the countries under analysis, containing indicators such as:

Independent Variables
The volumes of imports and exports
The volumes of the private investment and consumption
The volume of GDP
The GDP deflator
The government deficit as a percentage of the GDP
The sectorial value added
The general unemployment level and employment hours
The unit labour cost
Consumer surveys and economic sentiment indicators
Stock price and volatility indexes
Long-term interest rates

DATA: SAMPLE AND FREQUENCY

- **SAMPLE: 1998 TO 2014**

- **FREQUENCY: QUARTERLY**

- **SOME VARIABLES ARE MONTHLY (E.G. SURVEYS, STOCK MARKET VARIABLES, INTEREST RATES ETC.)**

- **WE ASSESS THREE WAYS TO COPE WITH THIS UNBALANCEDNESS PROBLEM.**

  1. USING THE QUARTERLY AVERAGES, I.E. AVERAGING OVER THE THREE MONTHS IN THE SAME QUARTER.

  2. USING THE LAST MONTH IN EACH QUARTER ONLY, IN ORDER TO USE THE MOST UPDATED INFORMATION AVAILABLE BUT AT THE COST OF A POSSIBLE INCREASE IN THE VOLATILITY OF THE VARIABLE.

  3. SPLITTING EACH MONTHLY VARIABLE INTO THREE QUARTERLY VARIABLES EACH CONTAINING, RESPECTIVELY, THE FIRST, SECOND AND THIRD MONTHS IN ALL QUARTERS ACROSS THE PERIOD. THIS APPROACH IS IN THE SPIRIT OF THE UMIDAS REGRESSIONS INTRODUCED IN FORONI, MARCELLINO SCHUMACHER (2014, JRSS). IT PRESERVES ALL THE AVAILABLE HIGHER FREQUENCY INFORMATION AND MODEL LINEARITY, AT THE COST OF FURTHER ENLARGING THE NUMBER OF VARIABLES UNDER CONSIDERATION

MODELS

- **WE ASSESS A LARGE SET OF SHORT TERM FORECASTING MODELS, FOCUSING ON THOSE THAT PERFORM WELL FOR MACROECONOMIC VARIABLES BASED ON PREVIOUS STUDIES.**

- **THE MODELS THAT WE CONSIDER ARE ALSO SIMILAR TO THOSE USED FOR THEIR SHORT TERM PROJECTIONS BY THE STATE PROJECTIONS CONSORTIUM (RESPONSIBLE FOR DEVELOPING EMPLOYMENT PROJECTIONS UNDER CONTRACT WITH THE U.S. DEPARTMENT OF LABOR).**

MODELS

- **FIVE MAIN CATEGORIES:**

    1. **AR(P)** WITH TWO LAG VALUES, **P = {1,4}**

    2. **ARMA(P,Q)** WITH **(P,Q) = (1,1)** AND **(P,Q) = (4,4)**

    3. **PC(F)** WITH **F = {1,2,3,4}** FACTORS AND INCLUDING OR EXCLUDING AN **AR(P)**

    4. **PLS(F)** WITH **F = {1,2,3,4}** FACTORS AND INCLUDING OR EXCLUDING AN **AR(P)**

    5. **BR(WN)** WITH SHRINKAGE PARAMETERS **W = {0.5,1,2}** AND INCLUDING OR EXCLUDING AN **AR(P)**

- **CHOICES FOR THE DETERMINISTIC COMPONENT:**

    1. THE CONSTANT AND SEASONAL DUMMY VARIABLES ARE NOT INCLUDED AT ALL

    2. ONLY THE CONSTANT IS INCLUDED,

    3. ONLY THE DUMMY VARIABLES ARE INCLUDED.

MODELS

■ **FORECAST COMBINATION SCHEMES ARE ALSO INCLUDED.**

✓ SA: Forecast Combination using a simple averaging scheme of all models

✓ Med: Forecast Combination using the median model

✓ TrA(j): Forecast Combination using a trimmed average of the best j = {2%,5%,10%,15%,20%,25%} models over a previous training sample

MODELS

- THERE IS ALSO A **SIMPLE LINEAR REGRESSION MODEL** WITH 5 REGRESSORS. THE REGRESSORS ARE DYNAMICALLY CHOSEN AT EACH POINT IN TIME AS THE 5 MOST CORRELATED EXPLANATORY VARIABLES WITH THE DEPENDENT VARIABLE (THE MODEL CONTAINS AN AR TERM TOO).

- THE REGRESSORS CAN BE ALSO USER DEFINED, E.G., TO ASSESS THE IMPACT OF SPECIFIC POLICY VARIABLES.

- WE DENOTE THIS MODEL BY SLR (WHICH STANDS FOR SIMPLE LINEAR REGRESSION)

FORECASTING EXERCISE

- ONCE THE MODEL PARAMETER ESTIMATES ARE OBTAINED, WE USE THE PROJECTION METHOD WHICH IS MORE ROBUST IN THE PRESENCE OF POSSIBLE MODEL MIS-SPECIFICATION (STOCK AND WATSON, 2002)

- WE START WITH $H = 1$ AND SET THE MAXIMUM FORECAST HORIZON AT $H = 8$. LONGER HORIZONS COULD BE ALSO CONSIDERED BUT TYPICALLY TIME SERIES BASED FORECASTS BECOME UNINFORMATIVE FOR LARGE VALUES OF H.

- THEN, WE SPECIFY THE EVALUATION PERIOD, *EVAL*, OMITTING THE FINAL H OBSERVATIONS IN ORDER TO HAVE THE SAME NUMBER (*EVAL*) OF FORECASTS FOR ANY GIVEN FORECAST HORIZON H, TO BE LATER COMPARED WITH THE ACTUAL VALUES.

Forecasting Exercise: Algorithm

1. Define the variable of interest, *Y*, where *Y* belongs to the set of European labour market variables

2. Use an initial sample of T_1 observations ($T_1 = T\text{-}Eval\text{-}h$).

3. Obtain the proper set of regressors for each model *I*, X_{IT}, $T = 1,2,\ldots, T_1$

4. For each forecast horizon regress Y_T on $X_{IT\text{-}H}$ and obtain the (method specific) vector of parameter estimates

5. Calculate for each method the forecasts using the projection method

6. Repeat the whole procedure increasing in each step the initial sample by one observation

7. Repeat the whole procedure for each variable of interest *Y*

FORECASTING EXERCISE: ALGORITHM

- FORECASTING SCHEMES:

    1. RECURSIVE (ADD ONE OBS AT THE END OF EACH SAMPLE, SO EXPANDING SAMPLE SIZE)

    ✓ More efficient

    2. ROLLING (DROP ONE OBS AT THE START AND ADD ONE AT THE END, SAME SAMPLE SIZE)

    ✓ More robust to in the presence of structural breaks

- IN OUR EXPERIMENTS:

    ✓ *h = 1,2,…,8*
    ✓ *Eval = 16* periods (2009-2012)
    ✓ CY, MT, LU and RO *Eval = 12* periods due to data limitations

EVALUATION CRITERIA – POINT FORECASTS

- **MEAN ABSOLUTE ERROR:** LESS SENSITIVE TO LARGE ERRORS, THAT ENTER LINEARLY RATHER THAN SQUARED

- **ROOT MEAN SQUARED FORECAST ERROR:** PERHAPS THE MOST COMMON CRITERION FOR FORECAST EVALUATION, AND IMPLIES THAT LARGE ERRORS ARE PROPORTIONALLY MORE IMPORTANT THAN SMALL ERRORS

- **SIGN SUCCESS RATIO:** PROVIDES A MEASURE OF DIRECTIONAL ACCURACY, WHICH CAN BE INTERESTING FOR EXAMPLE TO ASSESS WHETHER A GIVEN OCCUPATION OR SKILL WILL INCREASE OR DECREASE.

Interval Forecasts

- We also compute the **90%** and **95%** prediction intervals, to provide a measure of forecast uncertainty

- Based on either past forecast errors or bootstrapping

- We also include forecast interval evaluation measures (coverage rates)

DENSITY FORECASTS

- DENSITY FORECASTING IS BASED EITHER ON THE NORMALITY ASSUMPTION OR ON A MONTE CARLO SIMULATION METHOD, WHICH PROVIDES MORE COMPREHENSIVE EVALUATION OF FORECAST UNCERTAINTY.

- FOR EACH VARIABLE AND PROBABILITY LEVEL, WE CAN CONSTRUCT A FORECAST INTERVAL THAT CONTAINS THE FUTURE REALIZATION OF THE VARIABLE WITH THE CHOSEN PROBABILITY LEVEL.

- DENSITY EVALUATION IS THEN ALSO CONSIDERED (BERKOWITZ (2001) LIKELIHOOD RATIO STATISTICS).

EARLY WARNING SYSTEM (EWS)

- AN EARLY WARNING SYSTEM (EWS) IS CONSTRUCTED BASED ON THE 90% AND 95% PREDICTION AND DENSITY FORECASTING INTERVALS.

- THE EWS COMPARES THE HISTORICAL AVERAGE OF EACH VARIABLE WITH ITS BEST SHORT TERM FORECAST, TO HIGHLIGHT WHICH VARIABLES ARE MOST LIKELY TO INCREASE OR DECREASE MOST OVER THE FORECASTING PERIOD.

- THE EWS INDICATES WHICH OCCUPATIONAL TYPES WILL BE ON PARTICULARLY HIGH/LOW DEMAND OVER THE FORECAST PERIOD

RESULTS: ISCO (4034 SERIES)

- **THE FORECASTS ARE GENERALLY CLOSE TO THE ACTUALS**

- **THE RESULTS ARE RELIABLE FOR MOST EUROPEAN COUNTRIES, EVEN WHEN DISAGGREGATING BY AGE, SEX AND OCCUPATIONAL TYPE**

- **RMSFE: A TRIMMED (2%) MEAN IS ON AVERAGE THE BEST PERFORMER**

RESULTS: ISCO (4034 SERIES)

- MODELS BASED ON EXTERNAL MACRO INFORMATION ARE SELDOM BEST, JUST SOMETIMES WHEN LOOKING AT THE LARGEST FRACTION OF VARIABLES BASED ON THE SSR CRITERION

- THERE ARE ONLY MINOR DIFFERENCES BETWEEN ROLLING AND RECURSIVE ESTIMATION

- GOOD PERFORMANCE OF THE AR AND ARMA MODELS BUT ALSO INDICATE THAT MACROECONOMIC INFORMATION, AS SUMMARISED BY THE PC, PLS AND BR METHODS, IS RELEVANT FOR A NON-NEGLIGIBLE FRACTION OF VARIABLES
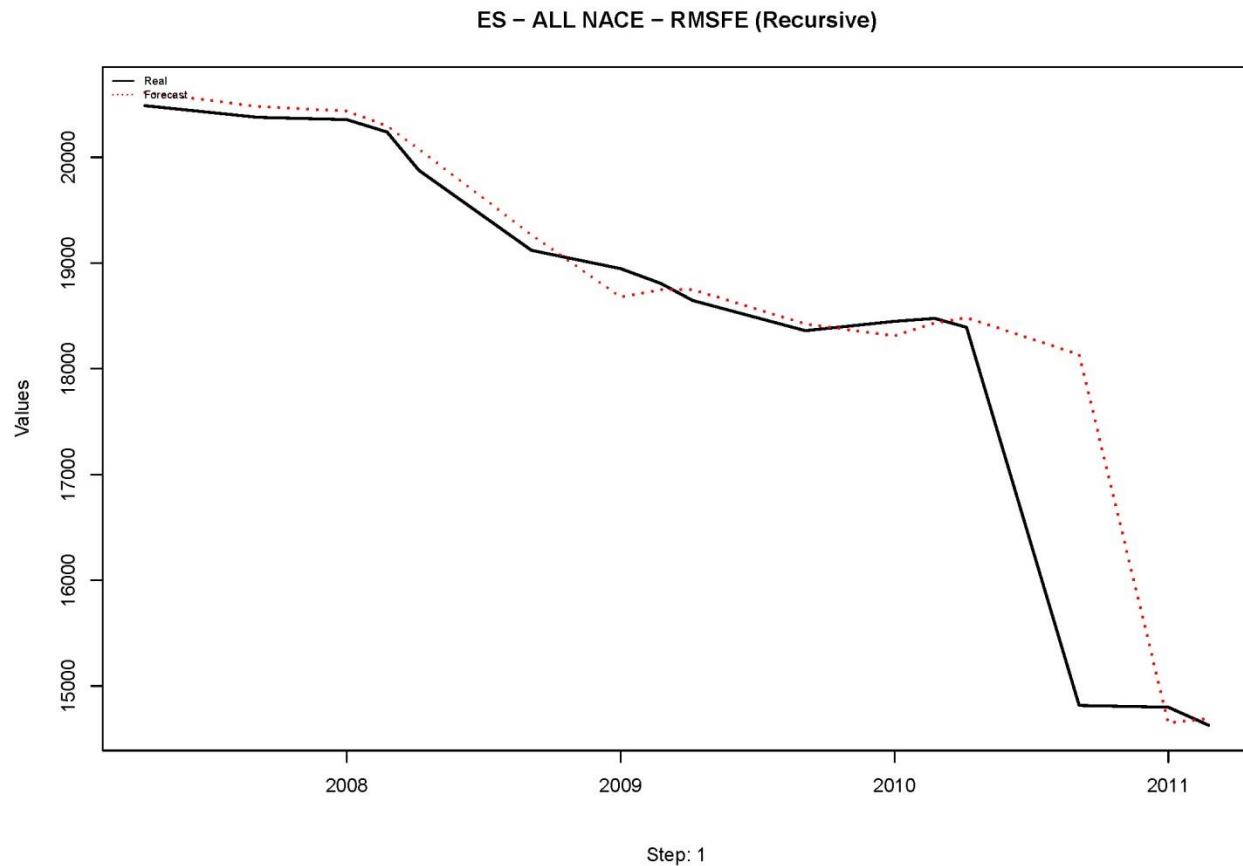
RESULTS: NACE (3586 SERIES)

- IN GENERAL, THE (OPTIMAL 1-STEP AHEAD) FORECASTS ARE QUITE CLOSE TO THE ACTUALS

- STANDARD ARMA MODELS IN GENERAL PERFORM VERY WELL WHEN H = 1,2

- HOWEVER, INTERESTINGLY, FOR LONGER FORECAST HORIZONS THE ROLE OF THE MACROECONOMIC INFORMATION INCREASES, WITH PC, PLS AND BR OFTEN PROVIDING THE BEST FORECAST IN TERMS OF RMSFE.

- THIS IS OVERALL IN LINE WITH ECONOMIC THEORY, WHICH PREDICTS THAT IT TAKES SOME TIME FOR MACROECONOMIC AND FINANCIAL SHOCKS TO MANIFEST THEIR EFFECTS ON LABOUR MARKET VARIABLES.
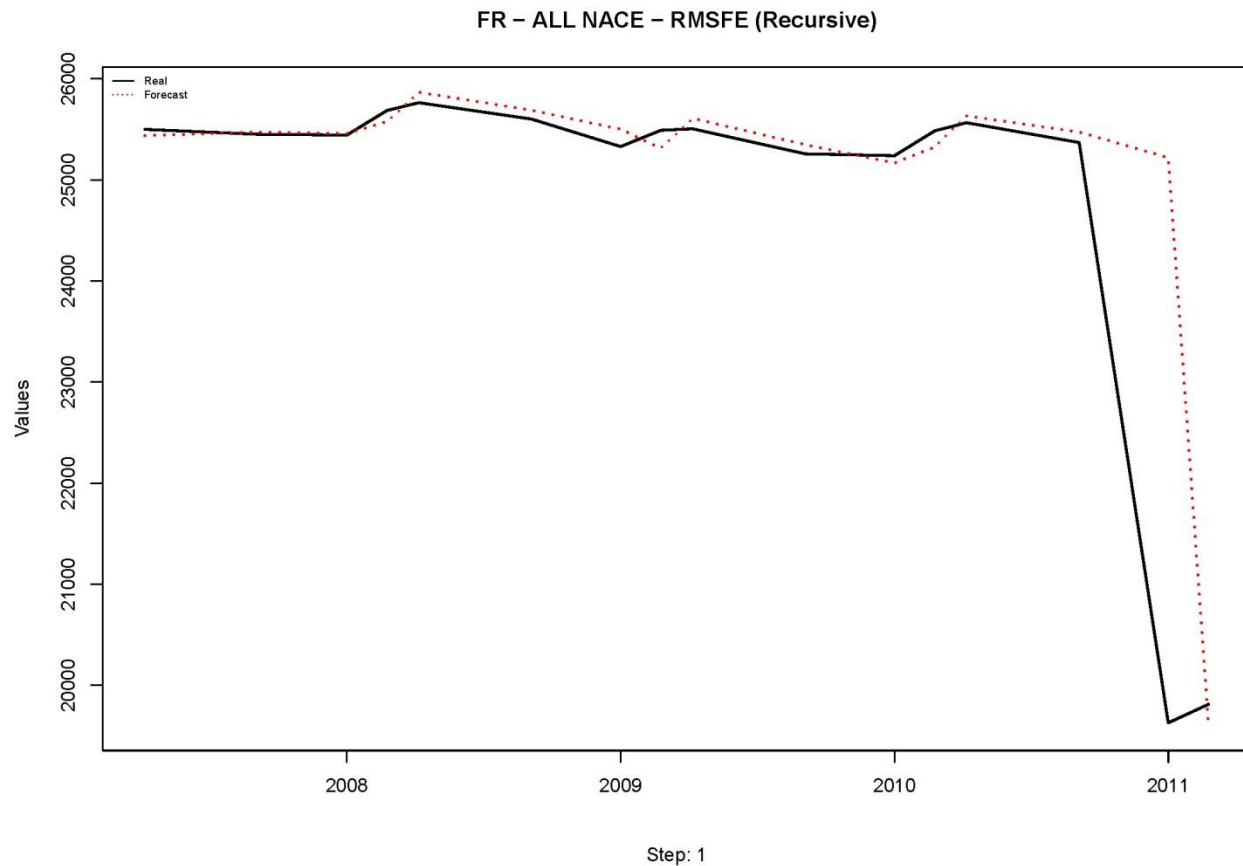
RESULTS: NACE

DE – ALL NACE – RMSFE (Recursive)

RESULTS: NACE

ES − ALL NACE − RMSFE (Recursive)
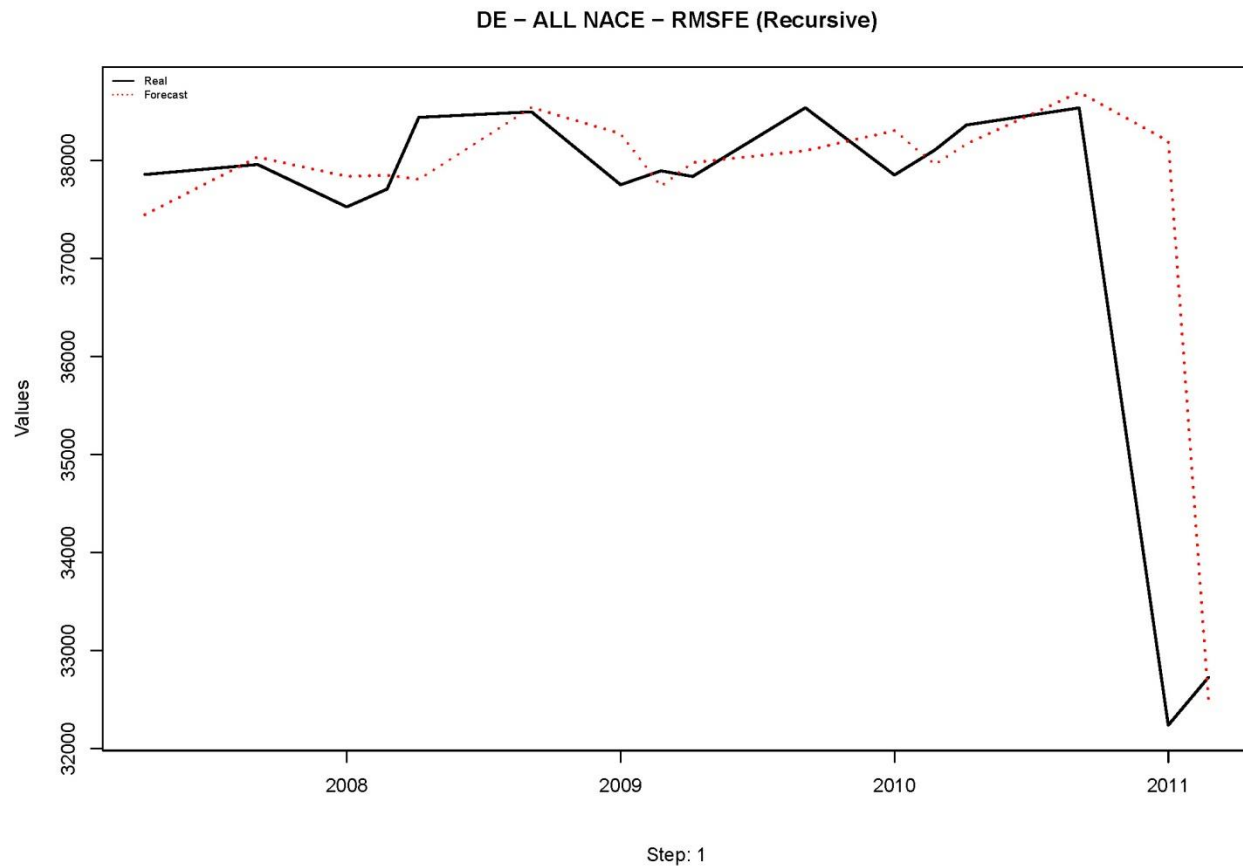
RESULTS: NACE

FR − ALL NACE − RMSFE (Recursive)

RESULTS: SKILLS (470 SERIES)

- IT AGAIN EMERGES THAT IN GENERAL THE (OPTIMAL 1-STEP AHEAD) FORECASTS ARE QUITE CLOSE TO THE ACTUALS.

- IN TERMS OF DISAGGREGATE RESULTS, WE HAVE SIMILAR COMMENTS AS FOR THE SECTORAL EMPLOYMENT CASE.

- IN PARTICULAR, ALSO FOR SKILLS STANDARD ARMA MODELS IN GENERAL PERFORM VERY WELL WHEN H = 1,2, WHILE FOR LONGER FORECAST HORIZONS THE ROLE OF THE MACROECONOMIC INFORMATION INCREASES, WITH PC, PLS AND BR OFTEN PROVIDING THE BEST FORECAST IN TERMS OF RMSFE.

COMPARISON WITH DG ECFIN FORECASTS

- AN INTERESTING COMPARISON IS WITH THE ECFIN FORECASTS OF TOTAL EMPLOYMENT FOR THE MAJOR EU ECONOMIES: DE, ES, FR AND IT.

- AT FIRST WE INVESTIGATE WHICH OF THE MODELS SEEMS TO PROVIDE THE MOST ACCURATE FORECASTS IN THE CROSS-VALIDATION STUDY. THEN, WE SELECT THIS MODEL AND WE COMPARE ITS FORECASTS TO THE CURRENT ECFIN FORECASTS. WE DO THIS IN GROWTH RATES RATHER THAN LEVELS OF THE SERIES IN ORDER TO ACCOMMODATE THE FACT THAT WE CONSTRUCTED "OUR" AGGREGATES USING THE SERIES WITH THE MAXIMUM AVAILABLE INFORMATION (SO THE TARGET VARIABLES ARE DIFFERENT FOR US AND FOR ECFIN).

- THE OBSERVATIONS FOR 2012 AND 2013 ARE ACTUAL VALUES, THOSE FROM 2014 TO 2015 ARE FORECASTS.

COMPARISON WITH DG ECFIN FORECASTS

- FOR GERMANY, FOR 2014 THE SBAS FORECASTS AGREE WITH THE OFFICIAL ONES AND SUGGEST A GROWTH BUT AT A HIGHER RATE OF 0.717% COMPARED TO 0.525% FOR ECFIN. THE GROWTH IS FORECASTED TO CONTINUE IN 2015, WHERE AGAIN THE SBAS SYSTEM SUGGESTS THAT THE EXPANSION WILL BE SHARPER, 1.799% COMPARED TO 0.642%.

- FOR SPAIN, BOTH FORECASTS ARE POSITIVE AND EQUAL TO 0.226% AND 0.6% FOR THE OFFICIAL AND SBAS RESPECTIVELY. FOR 2015 THE SBAS SUGGESTS THAT THERE WILL BE POSITIVE GROWTH OF 0.063%, LOWER THAN THE 0.861% OF THE OFFICIAL FORECAST.

COMPARISON WITH DG ECFIN FORECASTS

- FOR FRANCE, WE SEE DISCREPANCIES IN THE SIGNS, WITH THE OFFICIAL FORECAST BEING -1.080% FOR 2014 AND 1.105% FOR 2015. THE RELEVANT SBAS FORECASTS ARE 2.712% AND -4.096% RESPECTIVELY.

- FOR ITALY, BOTH FORECASTS AGREE ON THE SIGN FOR 2014. THE OFFICIAL FORECAST IS A CONTRACTION OF 0.229% WHEREAS THE SBAS SUGGESTS A CONTRACTION OF -0.86%. FOR 2015 THE SBAS SUGGESTS SLIGHTLY NEGATIVE GROWTH OF -0.002% WHICH IS IN CONTRAST TO THE OFFICIAL FORECAST WHICH SUGGESTS AN EXPANSION OF 0.474%.

AN EXAMPLE OF THE EWS

DE, ECOFIN, $h = 1$. EWS Example

Date	Value	EWSDens	Growth
2009.4	38555.71	**	
2010.1	37867.45	**	−1.80%
2010.2	38111.13	*	0.64%
2010.3	38392.53		0.74%
2010.4	38554.46	**	0.42%
2011.1	32199.27		−18.01%
2011.2	32734.43	**	1.65%
2011.3	32889.77	**	0.47%
2011.4	33100.31	*	0.64%
2012.1	32066.86		−3.17%
2012.2	32346.65		0.87%
2012.3	32757.32		1.26%
2012.4	32767.08		0.03%
2013.1	32290.68		−1.46%
2013.2	32812.66		1.60%
2013.3	33084.89		0.83%

CONCLUSIONS

- IN GENERAL OUR RESULTS INDICATE THAT FORECAST COMBINATION (A TRIMMED MEAN) WORKS BEST, WITH AR AND ARMA MODELS BEST IN TERMS OF THE FRACTION OF VARIABLES FOR WHICH THEY PRODUCE THE BEST FORECASTS, IN LINE WITH THE LITERATURE ON PREDICTING MACROECONOMIC VARIABLES

- MACROECONOMIC INFORMATION, COMBINED WITH VARIABLE REDUCTION METHODS, TENDS ONLY SOMETIMES TO PREDICT THE FUTURE DIRECTION OF THE SERIES (SIGN FORECASTING) MORE ACCURATELY ON AVERAGE

CONCLUSIONS

- HOWEVER, THE DISAGGREGATE RESULTS SHOW THAT MACROECONOMIC INFORMATION IS RELEVANT FOR A NON-NEGLIGIBLE FRACTION OF THE SERIES SINCE, IN PARTICULAR AT LONGER HORIZONS, PC, PLS OR BR MODELS ARE OFTEN THE BEST

- ALL THESE FINDINGS ARE QUITE ROBUST TO THE CHOICE OF THE LOSS FUNCTION (MAE, RMSFE OR SSR), ESTIMATION METHOD (ROLLING OR RECURSIVE) AND TO WHETHER GROUPING IS BASED ON EQUAL OR UNEQUAL WEIGHTING

CONCLUSIONS

- WE ALSO PRODUCE INTERVAL AND DENSITY FORECASTS FOR ALL VARIABLES, WHICH PROVIDES AN INDICATION OF THE UNCERTAINTY AROUND THE POINT FORECASTS.

- BASED ON THE INTERVAL AND DENSITY FORECASTS, WE HAVE ALSO DESIGNED AND IMPLEMENTED AN EARLY WARNING SYSTEM THAT HIGHLIGHTS WHICH EMPLOYMENT CATEGORIES AND TYPES AND MORE LIKELY TO BE VERY DIFFERENT IN THE FUTURE COMPARED WITH THE PAST

CONCLUSIONS

- OVERALL, WE OBTAIN A DETAILED AND RELIABLE SHORT TERM SECTORAL BASED ANTICIPATORY SYSTEM (SBAS) FOR THE EUROPEAN LABOUR MARKET, AND WE BELIEVE IT CAN BE USEFUL FOR THE DESIGN OF ECONOMIC, SOCIAL AND VOCATIONAL TRAINING POLICIES

- NATURALLY, THE PROTOTYPE SBAS STILL REQUIRES SOME REFINEMENTS AND IMPROVEMENTS BASED ON THE ANALYSIS OF THE RESULTS, AND ALSO ON THE REACTION OF CEFEFOP AND THE INPUT OF KEY STAKEHOLDERS AND EXPERTS, INCLUDING CEDEFOP'S SKILLNET.

- WE LOOK FORWARD TO WORKING ON THESE ADDITIONAL IMPROVEMENTS OF THE SBAS, WITH THE PRECIOUS INPUT FROM AND COLLABORATION WITH CEDEFOP AND RELEVANT STAKEHOLDERS.