

Forecasting with many predictors

Scientific Coordinator:
Prof. Moisă Altăr, Ph.D.

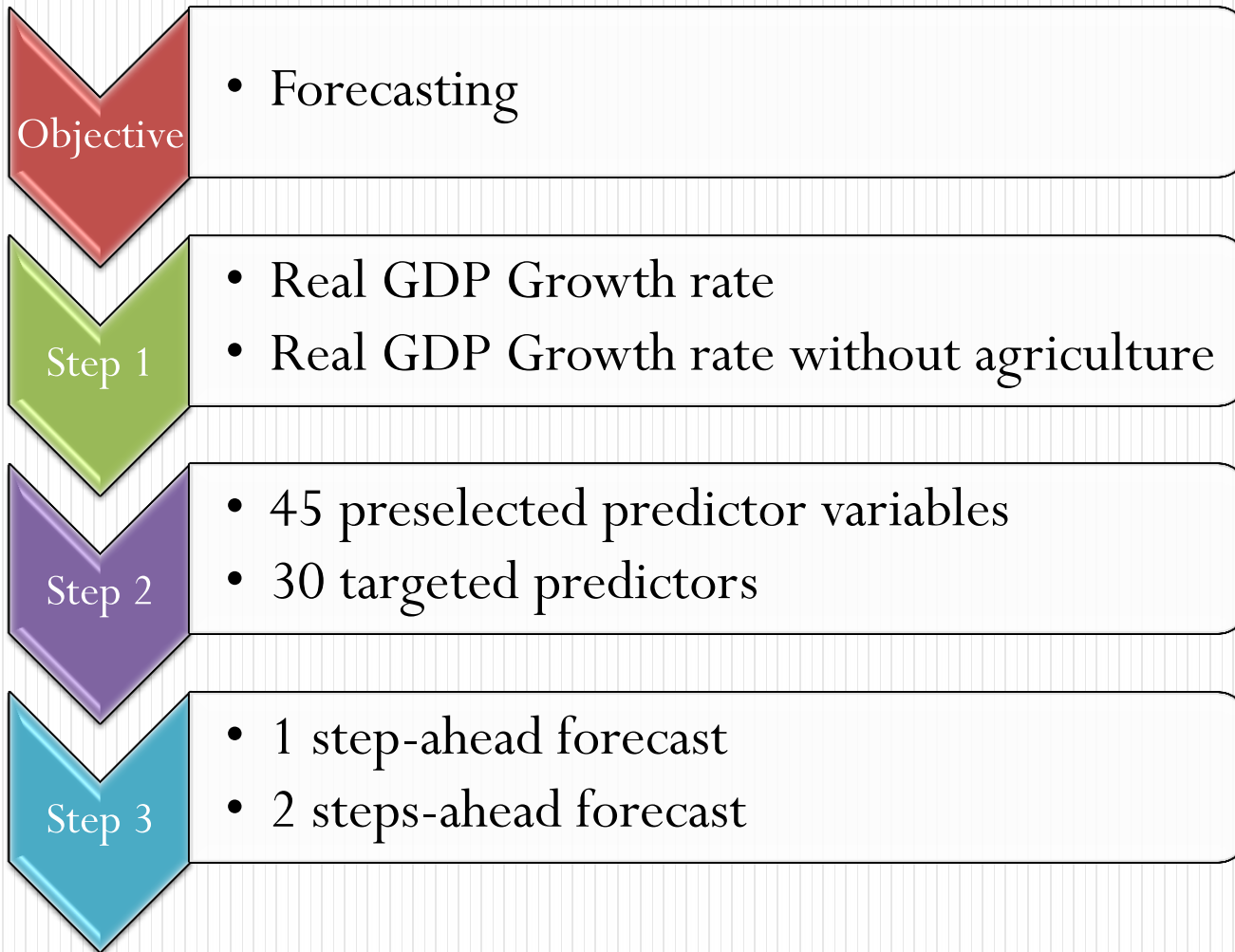
MSc. Student:
Marian-Alexandru Zechiu
DOFIN, 2014

Contents

- Objective
- Literature Review
- Methodology
- Data
- Results
- Conclusions
- Bibliography
- Appendix

Objective

- Short-term forecasting using a large number of predictor variables
- Important issues arise:
 1. How to use as much information as possible, without a big increase in the complexity of the model?
 2. Which predictors to use?
 3. Which predictors are best for forecasting?



Literature review

- *Sargent and Sims (1977)* - two factors could explain a large fraction of the variance of important U.S. quarterly macroeconomic variables
- *Stock and Watson (1998, 2002)* - made these types of forecasting more feasible
- *Boivin and Ng (2006)* - more is not always better
- *Bai and Ng (2008)* - targeted predictors
- *Eickmeier and Zeigler (2008)* - analysis performed over 52 studies related to forecasting output and/or inflation with large-scale factor models – SW most prevalent
- Other related work: *Forni et al (2000, 2005)*, *Kapetanios and Marcellinio (2006)*, *Doz et al (2011)*

Methodology

- Used Stock-Watson methodology to obtain factors and subsequent forecasts
- Given X_{it} an element of a $T \times N$ matrix X_t of predictors (N predictors, T time periods), we wish to find those factors F_t for which
- $X_{it} = \lambda_i \mathbf{x} F_t + e_{it}$, by minimizing MSE:

$$\min \frac{1}{N \times T} \times \sum_{i=1}^N \sum_{t=1}^T e_{it}^2$$

- Solution is F_t – the principal components formed w.r.t. X_t
- Select a subset of factors $f_t \subset F_t$ and use them in forecasting equation:
 - $y_{t+h} = \alpha(L) \times y_t + \beta(L) \times f_t + e_{t+h}$
- Only first factors are selected, those with economic meaning

Targeted Predictors

- Targeted predictor: variable which is tested to have predictive power over GDP
- Targeted predictors are formed using Bai-Ng (2008) methodology \rightarrow soft-thresholding
- Targeted predictors are computed using the Least Absolute Shrinkage and Selection Operator (LASSO)

LASSO

- Let SSR be the sum of squared residuals from regressing y_{t+h} on all X_{it}
- The LASSO is the solution to the optimization problem:

$$\begin{aligned} & \min \text{SSR} \\ & \sum_{i=0}^N |\beta_i| \leq \rho \end{aligned}$$

- ρ is a tuning parameter, which can take values $\in [0; \sum_{i=0}^N |\beta_i^{OLS}|]$

- ρ is used to control which β 's are 0 and which are not 0
- There is no exact solution for the LASSO
- Use a modification to the Least Angle Regression (LARS) algorithm which gives the full LASSO path
- Observation! We obtain a set of targeted predictors for each step of the analysis ($h=1,2$)

Data

- 45 predictor variables, from 8 groups:
 - Economic Sentiment Indicator (5)
 - GDP by Sector (10)
 - External Sector (6)
 - GDP by Components (5)
 - Labor Market (4)
 - Industrial Production (6)
 - Deflators (6)
 - Financial Indicators (3)
- ❖ For real GDP growth without agriculture, GDP by Sector will have 9 elements

- Final data (QoQ, 2003Q1 – 2013Q4) is
 - Seasonally-adjusted
 - Stationarized
 - Standardized
- PCA is performed on final data
- LASSO is performed on final data

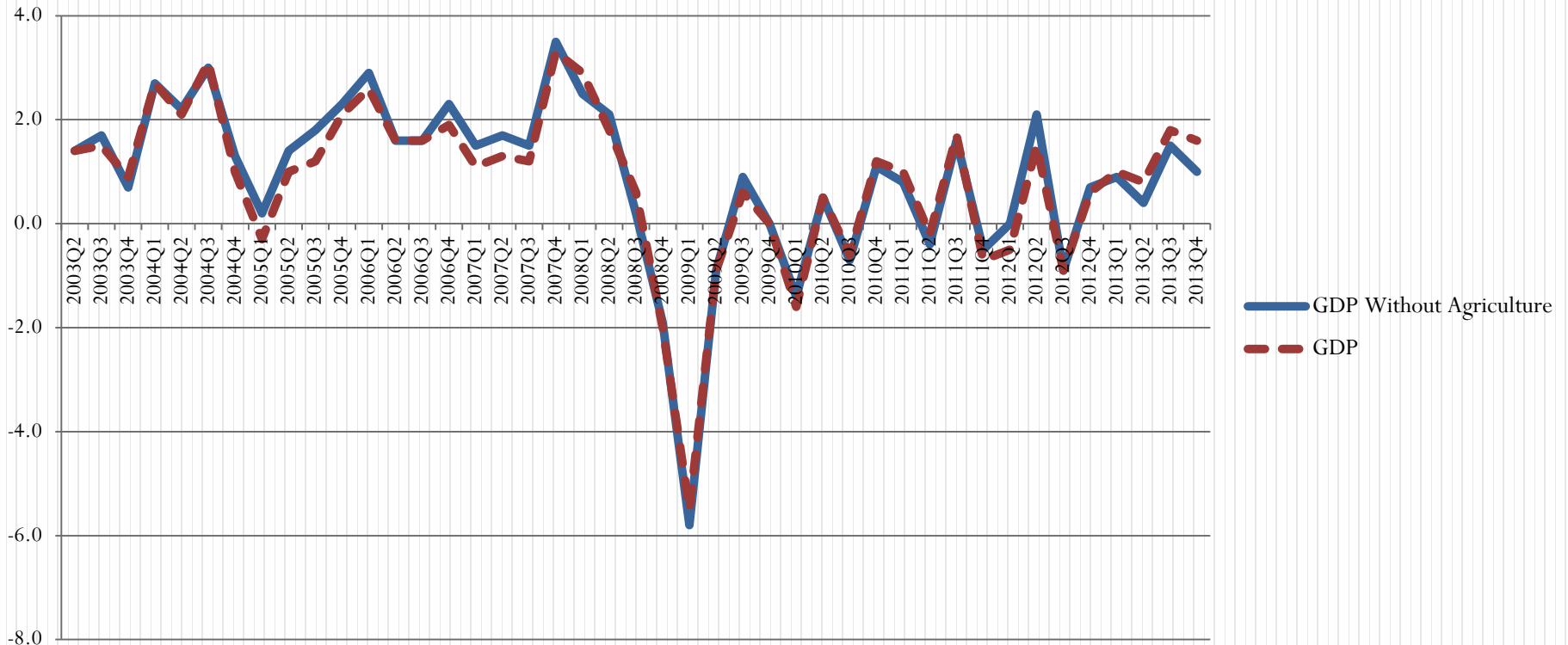


Fig. 1. GDP growth rate

Results

- After extracting factors, first step is factor identification.
- Compute coefficient of determination (R^2) between each component and each predictor
- Top coefficients identify the principal component

Factor Identification

- First factor is always
 - a measure of the Economic Sentiment Indicators (especially Industrial and Consumer Confidence)

GDP - First Factor					
	45 predictors		1 step targeted		2 steps targeted
Industrial confidence indicator	61.29%	Industrial confidence indicator	63.25%	EUR/RON	51.18%
Imports of goods and services	56.58%	Consumer confidence indicator	52.03%	Industrial confidence indicator	50.40%
Consumer confidence indicator	50.77%	Imports of goods and services	49.04%	Households final consumption - Deflator	48.12%
Households final consumption - Deflator	46.98%	Households final consumption - Deflator	46.61%	Consumer confidence indicator	46.06%

Table 1. R^2 (%) for F_1

GDP w/o Agriculture - First Factor					
	44 predictors		1 step targeted		2 steps targeted
Industrial confidence indicator	61.21%	Industrial confidence indicator	67.08%	Industrial confidence indicator	60.65%
Imports of goods and services	56.57%	Imports of goods and services	54.21%	EUR/RON	52.01%
Consumer confidence indicator	50.70%	Consumer confidence indicator	52.81%	Households final consumption - Deflator	51.97%
Households final consumption - Deflator	47.03%	Retail confidence indicator	41.92%	Consumer confidence indicator	47.31%

Table 2. R^2 (%) for F_1

- The preselected variables - “an informative guess”
- The above mentioned statement holds even for targeted predictors
- The Economic Sentiment Indicators have an important role to play in GDP forecasting

- Second factor is either:
 - a measure of economic output (industrial production and final consumption)
 - ✓ 1 step ahead targeted predictors and all predictor models
 - a financial related factor (interest rate)
 - ✓ 2 steps ahead targeted predictors

GDP - Second Factor					
	45 predictors		1 step targeted		2 steps targeted
Households final consumption	48.20%	Households final consumption	49.54%	12-month money market interest rate	69.62%
Industry	42.68%	Industrial Production of Capital goods	39.65%	Average Policy rate	35.65%
Industrial Production of Capital goods	33.13%	Labor cost index	32.85%	Labor cost index	27.34%
Gross domestic product - deflator	25.44%	Euro area GDP growth (18 countries)	23.18%	Industrial Production of Energy	25.26%

Table 3. R^2 (%) for F_2

GDP w/o Agriculture - Second Factor					
	44 predictors		1 step targeted		2 steps targeted
Households final consumption	46.24%	Households final consumption	49.69%	12-month money market interest rate	60.60%
Industry	39.22%	12-month money market interest rate	36.23%	Households final consumption	44.15%
Industrial Production of Capital goods	29.79%	Industrial Production of Capital goods	32.90%	Labor cost index	32.22%
Exports of goods and services - deflator	25.29%	Labor cost index	31.43%	Industry	32.06%

Table 4. R^2 (%) for F_2

- Industrial sector holds a total of approximately 42% of GDP (Source: The World Bank)
- Final consumption of households holds a total of approximately 72% of total GDP (Source: The World Bank)

- Third factor is most likely:
 - a measure of economic output (GVA in Agriculture, or final consumption)
 - a labor market measure (Labor Productivity or Employment rate)
 - a financial related factor (interest rate or exchange rate)

GDP - Third Factor					
	45 predictors		1 step targeted		2 steps targeted
12-month money market interest rate	56.36%	Agriculture, forestry and fishing	37.22%	Households final consumption	39.37%
Average Policy rate	45.48%	Gross capital formation	29.99%	Gross capital formation	24.88%
Agriculture, forestry and fishing	32.07%	12-month money market interest rate	26.88%	Exports of goods and services	17.87%
Public administration, education, human health	22.39%	Industrial Production of Capital goods	23.48%	Agriculture, forestry and fishing	17.06%

Table 5. R^2 (%) for F_3

GDP w/o Agriculture- Third Factor

	44 predictors		1 step targeted		2 steps targeted
12-month money market interest rate	57.31%	Hourly labor productivity	38.26%	Information and communication	33.80%
Average Policy rate	41.65%	Employment rate	35.29%	Services Confidence Indicator	27.50%
Financial and insurance activities	22.77%	Gross capital formation - deflator	27.51%	Employment rate	25.45%
EUR/RON	22.08%	Current account, Income	22.64%	EUR/RON	22.96%

Table 6. R^2 (%) for F_3

- The third factor contains useful information about the labor market, which makes it different from the previous ones
- Labor Productivity has a very high correlation with this third factor

Power of prediction

First 10 Targeted Predictors for GDP		
Crt. Nr.	1-step-ahead	2-steps-ahead
1	Imports of goods and services	Current account, Income ✓
2	Consumer confidence indicator	Industrial production of durable consumer goods ✓
3	Financial and insurance activities ✓	Wholesale and retail trade, transport, accommodation and food service activities
4	Agriculture, forestry and fishing	Information and communication
5	Labor cost index ✓	Current account, Current transfers
6	Industrial production of durable consumer goods ✓	Current account, Services
7	Capital account ✓	Financial and insurance activities ✓
8	Construction	Labor cost index ✓
9	Current account, Income ✓	Capital account ✓
10	Gross capital formation - Deflator	Exports of goods and services

Table 7. Targeted predictors for GDP

- The LASSO yields that some variables have improved predictive power over both 1 and 2 steps-ahead GDP forecast (5 variables)

Power of prediction – Discussion I

- Consumer Confidence Indicator has good predictive power for the 1-step-ahead GDP and for the 1-step-ahead and 2-steps-ahead GDP w\o Agriculture forecasts, consistent with previous findings of factor identification
- Labor Cost Index (total financial burden arising from the use of the labor production factor) also has great predictive power over both GDP and GDP w\o Agriculture series & can be easily be controlled by the Central Government

First 10 Targeted Predictors for GDP w/o Agriculture		
Crt. Nr.	1-step-ahead	2-steps-ahead
1	Consumer confidence indicator ✓	Industry
2	Financial and insurance activities ✓	Current account, Income
3	Labor cost index ✓	Wholesale and retail trade, transport, accommodation and food service activities
4	Industrial production of durable consumer goods ✓	Information and communication
5	Construction	Industrial production of durable consumer goods ✓
6	Capital account ✓	Labor cost index ✓
7	12-month money market interest rate	Financial and insurance activities ✓
8	Current account, Income	Consumer confidence indicator ✓
9	Gross capital formation - deflator	Capital account ✓
10	Households' final consumption expenditures ✓	Households' final consumption expenditures ✓

Table 8. Targeted predictors for GDP w/o Agriculture

- The LASSO yields that some variables have improved predictive power over both 1 and 2 steps-ahead GDP w/o Agriculture forecast (6 variables)

Power of prediction – Discussion II

- Industrial production of durable consumer goods (e.g. cars) is a key variable for both GDP and GDP w\o Agriculture analyses
- Financial and Insurance Activities are more important in predicting GDP without Agriculture growth (2nd place) than real GDP growth (3rd place)

Explained Variance

GDP

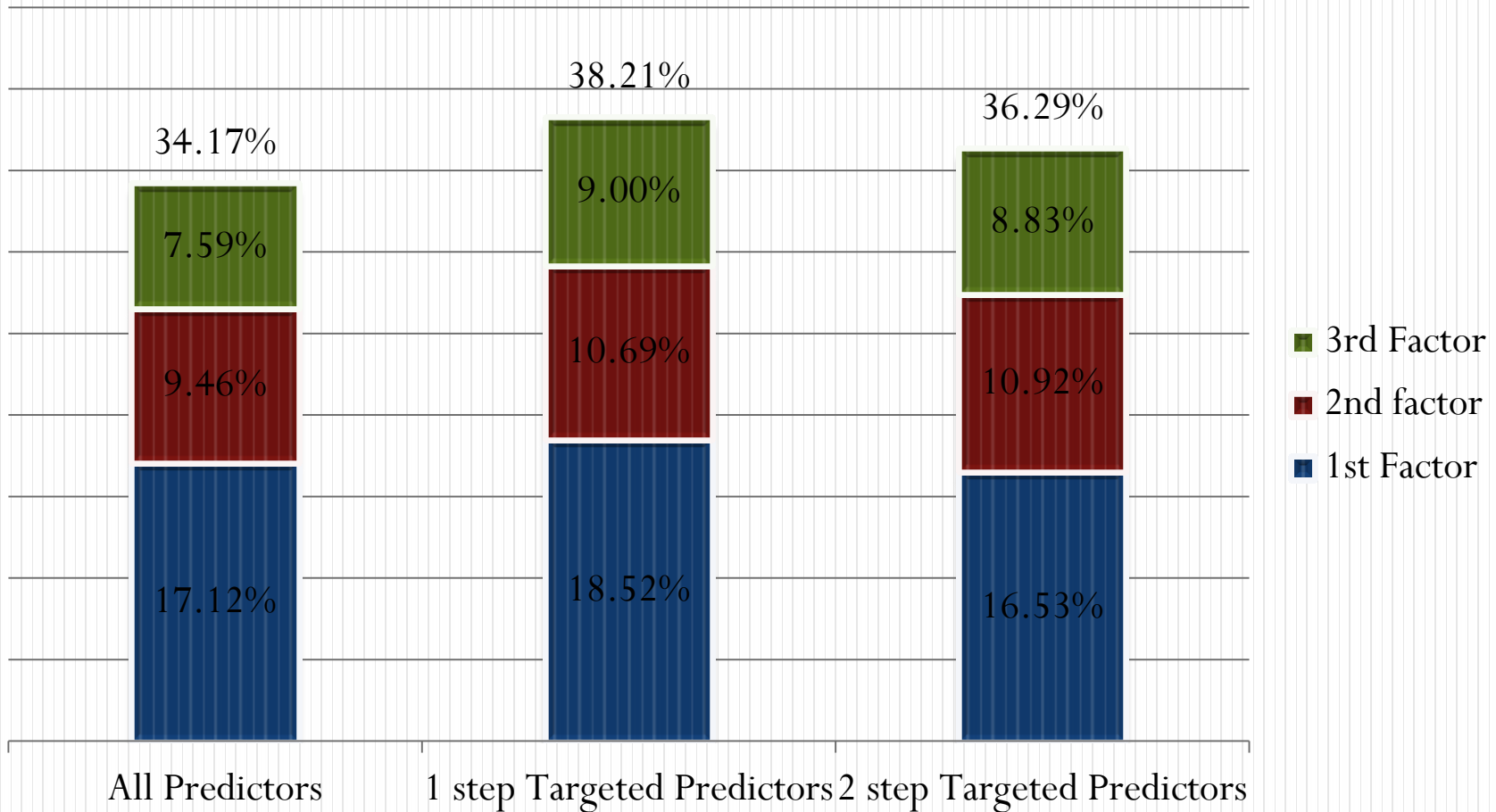


Fig. 2. Explained variance for real GDP growth

GDP w/o Agriculture

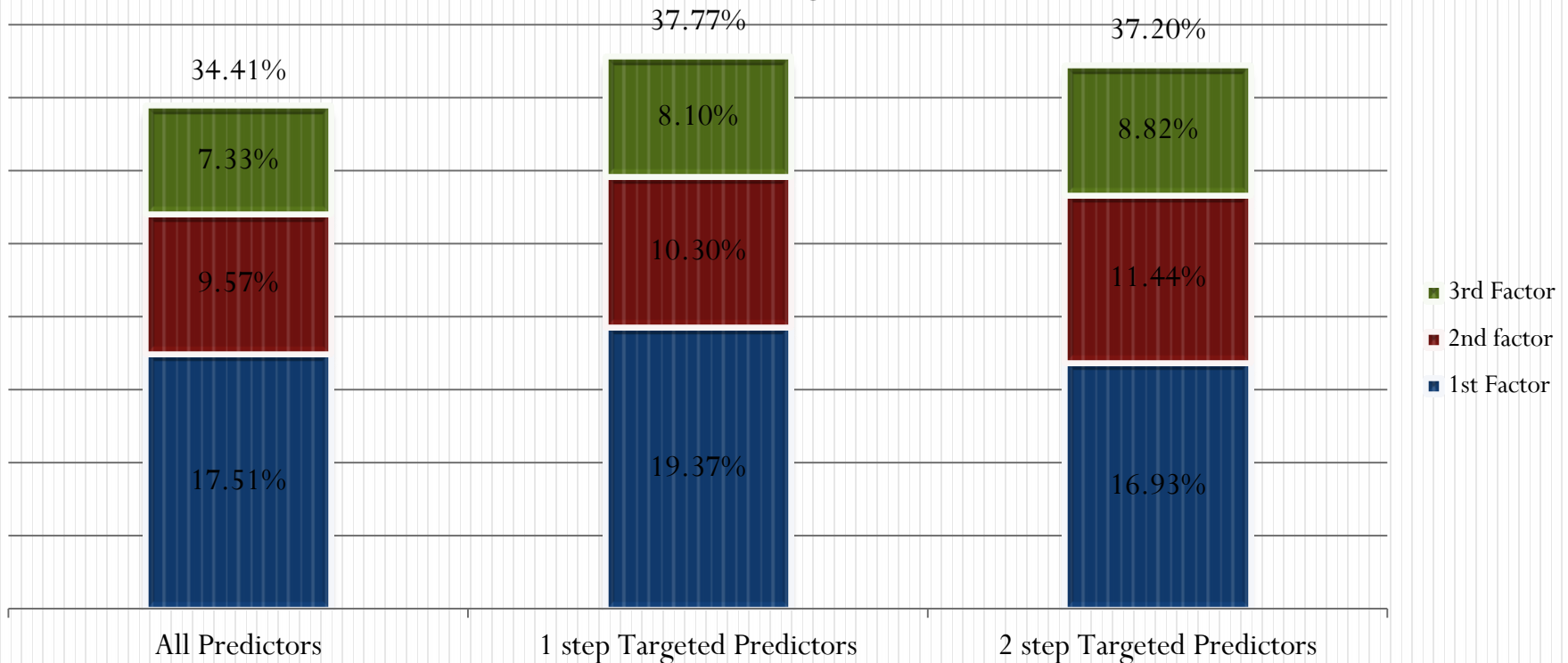


Fig. 3. Explained variance for real GDP growth w/o Agriculture

- In both analysis, targeted predictors explain a greater part of the information (which is, also, qualitatively better)

Further Analysis

- After factor identification, we perform regression analysis in the forecasting equation
 - In-sample 2003Q1 – 2011Q3 (80%)
 - Out-of-Sample 2011Q4 – 2013Q4 (20%)
- Optimal model is chosen w.r.t. 3 criteria:
 - Bayesian Information Criteria (BIC)
 - Root mean square forecast error (RMSFE)
 - Theil's U statistic (Theil U)
 - ❖ Special attention is devoted to Theil's U

Optimal model for GDP forecast						
		Theil U	RMSFE	BIC	Autoregressive component?	Factors included
Autoregressive	h=1	0.62	0.53	2.75	YES (1)	0
All Predictors	h=1	0.46	0.44	2.75	NO	2
	h=2	0.75	0.83	2.89	NO	3
Targeted Predictors	h=1	0.44	0.44	2.8	YES (1)	2
	h=2	0.78	0.66	2.86	NO	2

Table 9. Forecast Statistics

- Although not strictly dominant, *Targeted predictors' analysis* outperforms *All predictors' analysis* and *Autoregressive*
- In terms of Theil's U
 - *targeted predictors* outperform the *autoregressive analysis* by approximately 30%
 - *all predictors model* outperforms the benchmark by approximately 25%

Forecast Analysis – Discussion I

- Stock and Watson (*Stock and Watson (2002)*) obtained a 33% improvement in terms of MSFE over the benchmark
- Here, both *all predictors' model* and *targeted predictors* outperform benchmark by approximately 17% in terms of RMSFE
- Between the two types of predictor models:
 - For 1-step-ahead, there is an insignificant improvement of approximately 5% in terms of Theil's U for the *targeted predictors*, while the RMSFE performance is tied
 - For the 2-steps-ahead forecast, although Theil's U statistic is 4% worse for the *Targeted predictors*, the RMSFE improvement is of over 20%

Optimal model for GDP w/o Agriculture forecast						
		Theil U	RMSFE	BIC	Autoregressive component?	Factors included
Autoregressive	h=1	0.65	0.51	2.76	YES (1)	0
All Predictors	h=1	0.48	0.43	2.74	NO	2
	h=2	0.71	0.56	3.04	NO	2
Targeted Predictors	h=1	0.46	0.45	2.73	YES (1)	2
	h=2	0.72	0.59	2.95	NO	2

Table 10. Forecast Statistics

- Although not strictly dominant, *Targeted predictors' analysis* outperforms both *All predictors' analysis* and *Autoregressive* for 1-step-ahead forecast
- In terms of Theil's U:
 - *targeted predictors* outperform the *autoregressive analysis* by approximately 30%
 - *all predictors model* outperforms the benchmark by approximately 25%

Forecast Analysis – Discussion II

- In terms of RMSFE
 - *targeted predictors* outperform the *autoregressive analysis* by approximately 11%
 - *all predictors model* outperforms the benchmark by approximately 15%
- Between the two types of predictor models:
 - For 1-step-ahead, there is an insignificant improvement of approximately 4% in terms of Theil's U for the *targeted predictors*, while the RMSFE is 4% worse.
 - ✓ BIC statistic for *targeted predictors* is slightly better.
- Out-of-sample only has 9 observations

GDP w\o Agriculture

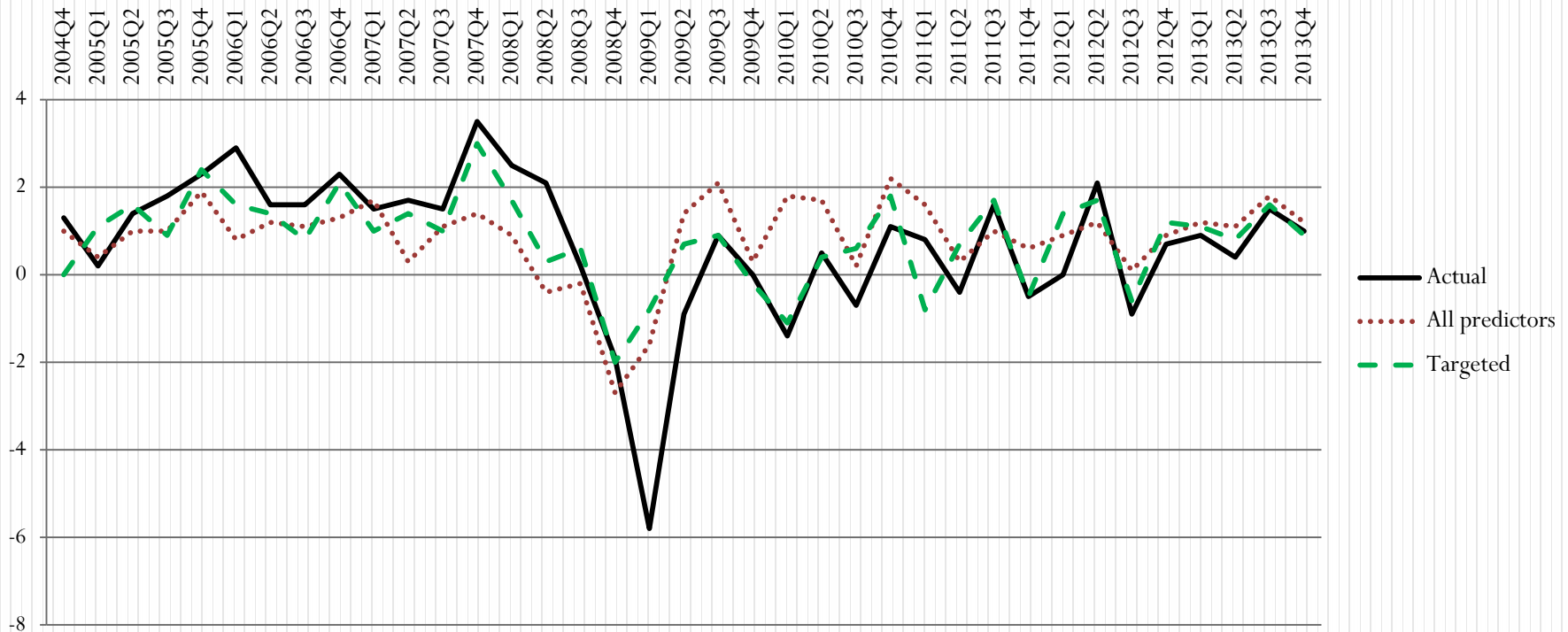


Fig. 4. GDP w/o Agriculture Graph

- For the GDP analysis, the optimal model with *Targeted predictors* obtains a 20% improvement in terms of RMSFE compared to the *All predictor model*
- For the GDP without agriculture analysis, the overall improvement is of approximately 15%, which can be viewed as an advantage towards using *targeted predictors* instead of the *all predictors' model*

Conclusions

1. Models including 2 factors generally perform the best: from a total of 8 optimal models, 7 of them had only two factors as exogenous variables for GDP forecast
2. The first 3 factors are, generally, well defined, with special emphasis being given on the 1st one
3. Three of the factors explain between 34% and 38% of the total information contained by the selected predictors
4. Targeted predictors improve overall forecasting performance, which is similar with findings in related literature (*Bai and Ng (2008)*)
5. A good way of improving the targeted predictor forecasts may be to give the inputs different weights in the principal components' analysis, weights which may take into account their predictive power as indicated by the LARS-LASSO algorithm

Thank you !

Bibliography

- Stock, J; Watson, M.; “*Macroeconomic Forecasting using Diffusion Indexes*” (2002), Journal of Business & Economic Statistics
- Bai, J; Ng, S; “*Forecasting economic time series using targeted predictors*” (2008), Journal of Econometrics
- Stock, J; Watson, M.; “*Dynamic Factor Models*” (2010), Oxford Handbook of Economic Forecasting
- Efron, B.; Hastie, T.; Johnstone, I; Tibshirani, R; “*Least Angle Regression*”, (2004), The Annals of Statistics

Appendix 1

LARS implementation for LASSO

1. Start with $r = y - \bar{y}$ ($\bar{y} = 0$)
2. Find predictor x_j most correlated with r
3. Move β_j from 0 $\rightarrow \beta_j^{OLS}$, until some other competitor x_k has as much correlation with the current residual as does x_j
4. Move β_j from last value to new β_j^{OLS} and β_k from 0 \rightarrow new β_k^{OLS} , until some other competitor x_l has as much correlation with the current residual.
 - New β_j^{OLS} and new β_k^{OLS} are defined by their joint least squares estimates

5. If a non-zero coefficient hits zero, drop its variable from the active set of variables and recompute the current joint least squares direction.
6. Continue in this way until all N predictors have been entered.

Appendix 2

Full data

Group	Name	Integration Level	Unit of Measurement
GDP by components	Gross domestic product at market prices less Agriculture	I(0)	Percentage change on previous period
	Gross domestic product at market prices	I(0)	Percentage change on previous period
	Final consumption expenditure of households	I(1)	Percentage change on previous period
	Final consumption expenditure of general government	I(0)	Percentage change on previous period
	Gross capital formation	I(1)	Percentage change on previous period
	Exports of goods and services	I(0)	Percentage change on previous period
	Imports of goods and services	I(0)	Percentage change on previous period
	Euro area real GDP growth	I(1)	Percentage change on previous period
External sector	Current account, Goods	I(1)	Net Flow Percentage of GDP
	Current account, Services	I(1)	Net Flow Percentage of GDP
	Current account, Income	I(1)	Net Flow Percentage of GDP
	Current account, Current transfers	I(1)	Net Flow Percentage of GDP
	Capital account	I(0)	Net Flow Percentage of GDP

Group	Name	Integration Level	Unit of Measurement
Economic Sentiment Indicator	Construction confidence indicator	I(1)	%
	Industrial confidence indicator	I(1)	%
	Retail confidence indicator	I(1)	%
	Consumer confidence indicator	I(1)	%
	Services Confidence Indicator	I(1)	%
Industrial Production	Current level of capacity utilization (%)	I(1)	%
	MIG - Intermediate goods	I(0)	Percentage change on previous period
	MIG - Energy (except Section E)	I(0)	Percentage change on previous period
	MIG - Capital goods	I(0)	Percentage change on previous period
	MIG - Durable consumer goods	I(0)	Percentage change on previous period
	MIG - Non-durable consumer goods	I(0)	Percentage change on previous period

Group	Name	Integration Level	Unit of Measurement
Labor Market	Employment rates by sex, age and nationality	I(1)	% (From 15 to 64 years)
	Unemployment rates by sex, age and nationality	I(1)	% (From 15 to 64 years)
	Labor cost index - Business economy (Industry, Financial & insurance services, Wholesale and retail trade, Transport services, IT&C, Professional, scientific and technical activities)	I(0)	Percentage change on previous period
	Hourly labour productivity	I(1)	Percentage change on previous period
Deflators	Gross domestic product at market prices	I(0)	Percentage change on previous period
	Final consumption expenditure of households	I(0)	Percentage change on previous period
	Final consumption expenditure of general government	I(0)	Percentage change on previous period
	Gross capital formation	I(0)	Percentage change on previous period
	Exports of goods and services	I(0)	Percentage change on previous period
	Imports of goods and services	I(0)	Percentage change on previous period

Group	Name	Integration Level	Unit of Measurement
Financial Indicators	EUR/RON	I(0)	Percentage change on previous period
	Average Policy rate	I(1)	% per year
	12-month money market interest rate	I(1)	% per year
GDP by Sector	Agriculture, forestry and fishing	I(0)	Percentage change on previous period
	Industry (except construction)	I(0)	Percentage change on previous period
	Construction	I(1)	Percentage change on previous period
	Wholesale and retail trade, transport, accommodation and food service activities	I(1)	Percentage change on previous period
	Information and communication	I(1)	Percentage change on previous period
	Financial and insurance activities	I(0)	Percentage change on previous period
	Real estate activities	I(1)	Percentage change on previous period
	Professional, scientific and technical activities; administrative and support service activities	I(0)	Percentage change on previous period
	Public administration, defence, education, human health and social work activities	I(0)	Percentage change on previous period
	Arts, entertainment and recreation; other service activities; activities of household and extra-territorial organizations and bodies	I(1)	Percentage change on previous period