EUROPEAN CENTRAL BANK

EUROSYSTEM

Variable selection for competitiveness analysis: What can we learn from LASSO and BMA?

B. Bluhm

15 December 2014

Outline of the presentation

1. Introduction
2. Review of BMA-results
3. Least absolute shrinkage and selection operator (LASSO)
   1. Methodology
   2. Literature
   3. Results
4. Conclusions

1. Introduction

- Research objective:
  - Identify most important drivers of export market shares in EU using quantitative model selection procedures
- Methodological approach:
  - Least absolute shrinkage and selection operator (LASSO)
- Dataset:
  - BMA-dataset (see Benkovskis/Bluhm/Bobeica/Osbat/Zeugner, 2014)

1. Introduction

- Motivation: Assess robustness of BMA-results using alternative model selection methods (Lisbon-workshop)
  - What are the most important drivers according to LASSO method?
  - Is there a common set of important drivers identified by different methods (LASSO and BMA)?
  - How stable are the estimated coefficient signs across methods?
  - How well can the models explain export market share dynamics in individual countries?

1. Introduction

- Summary of key results from LASSO
  - BMA-results remain very robust in several dimensions
  - Large overlap in set of most important drivers
  - Coefficient signs are the same across LASSO and BMA
  - LASSO tends to somewhat better explain export share dynamics in individual countries

2. Review of BMA-results

- Problem of model uncertainty: which subset of a potentially large set of drivers should be included in a model for competitiveness assessment?
- BMA-approach: estimate models with all possible combinations of variables
- For each variable, we get a probability to be included in the true model (posterior inclusion probability)
- We require two criteria to confirm the statistical significance of a given variable:
  - Posterior inclusion probability > prior inclusion probability
  - Abs(Posterior mean/posterior standard deviation) > 1.28

Model specification

- Static panel model with country fixed-effects
- 25 EU-countries, split in two samples of new and old member states
- Dependent variable: export market share growth
- 45 independent variables, 9 CompNet indicators
- Estimation period: 2002-2011 (independent variables), 2003-2012 (dependent variable)

2. Review of BMA-results: old EU member states

Explanatory Variables	BMA			
	PIP	Post. Mean	Post. SD.	Mean\SD.
New overlap with China	0,74	-0,33	0,13	-2,54
Crowding-out from China	0,72	-0,36	0,16	-2,25
Debt (% of total liabilities)	0,57	-0,19	0,09	-2,11
Investment (% of GDP)	0,50	0,31	0,19	1,63
Loans growth	0,42	0,15	0,09	1,67
Change in GVC position	0,41	-0,16	0,1	-1,60
Existent overlap with China	0,36	-0,44	0,35	-1,26
Labour productivity growth	0,34	0,16	0,12	1,33
Freedom to trade	0,33	0,20	0,14	1,43
TFP growth	0,33	0,22	0,2	1,10
Regulatory quality	0,33	0,13	0,09	1,44
Private consumption (% of GDP)	0,31	0,14	0,1	1,40
GVC participation	0,30	0,22	0,19	1,16
Public debt (% of GDP)	0,27	0,22	0,24	0,92
Legal system and property rights	0,26	0,11	0,09	1,22
Energy imports (% of energy use)	0,25	0,09	0,08	1,13
Size of government	0,25	-0,11	0,09	-1,22
Loans from foreign banks growth	0,25	-0,12	0,1	-1,20
Labour force with secondary ed.	0,24	0,10	0,09	1,11
HCI-CPI	0,24	-0,10	0,09	-1,11
R-squared	**0.383**			

2. Review of BMA-results: new EU member states

Explanatory Variables	BMA			
	PIP	Post. Mean	Post. SD.	Mean\SD.
Labour productivity growth	0,72	0,31	0,10	3,10
Existent overlap with China	0,70	-0,37	0,12	-3,08
Real FDI liabilities growth	0,28	0,22	0,10	2,20
RCA in high-tech ind.	0,27	-0,22	0,10	-2,20
TFP growth	0,21	0,28	0,15	1,87
Relative export prices adj. for quality	0,12	0,13	0,08	1,63
Labour force participation rate	0,11	-0,14	0,09	-1,56
New overlap with China	0,10	0,18	0,15	1,20
Share of construction in investment	0,10	0,16	0,12	1,33
Tax burden	0,08	0,11	0,08	1,38
Loans from foreign banks growth	0,08	-0,14	0,11	-1,27
Labour force with tertiary ed.	0,08	-0,13	0,19	-0,68
Size of government	0,07	-0,11	0,10	-1,10
GVC position	0,07	0,10	0,09	1,11
Patent applications	0,07	-0,10	0,11	-0,91
Legal system and property rights	0,07	0,09	0,09	1,00
Growth surprise	0,06	0,04	0,20	0,20
Investment growth	0,06	0,09	0,14	0,64
Freedom to trade	0,05	0,07	0,10	0,70
Loans (% of total liabilities)	0,05	-0,10	0,14	-0,71
R-squared	**0.408**			

3.1. LASSO: Methodology

- LASSO: Least absolute shrinkage and selection operator (Tibshirani, 1996)
- LASSO provides an efficient procedure to identify among large set of predictors the ones, which exhibit strongest effects
- LASSO is a combination of ridge regression and subset selection:
  - It shrinks some coefficients
  - It sets other coefficients to zero
- Why Shrinkage?
  - Prediction accuracy can sometimes be improved over OLS by shrinking or setting to 0 some coefficients
- When dealing with a large number of potential predictors, the fact that it sets coefficients to zero can be a big advantage for the sake of interpretation

3.1. LASSO: Methodology

Given a set of candidate drivers and a dependent variable y, LASSO fits a linear model:

(1) $$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + .. + \beta_k x_{ik} + \epsilon_i$$

The LASSO estimate is defined as:

(2) $$\beta^{LASSO} = \underset{\beta \epsilon \mathbb{R}^k}{argmin} \sum_{i=1}^{N} \varepsilon_i^2 + s \sum_{j=1}^{k} \left| \beta_j \right|$$

- Tuning parameter "s" controls the strength of the penalty. For s=0 we get $\beta^{LASSO} = \beta^{OLS}$, and $\beta^{LASSO} = 0$ when $s = \infty$.
- For values of s (s>=0) the solutions are shrunken versions of the least squares estimates. The penalty term causes some of the values to be shrunken to zero exactly

3.1. LASSO: Methodology

- The main difference between LASSO and other shrinkage methods (e.g. BMA, ridge regression) is that the latter only shrink coefficients towards zero but never sets them to zero exactly
- LASSO performs shrinkage and variable selection simultaneously
- The computation of the lasso solution is a quadratic programming problem, and can be tackled by standard numerical algorithms
- Least angle regression (LARS) procedure provides an efficient way to compute the solutions to the LASSO problem

LARS algorithm:

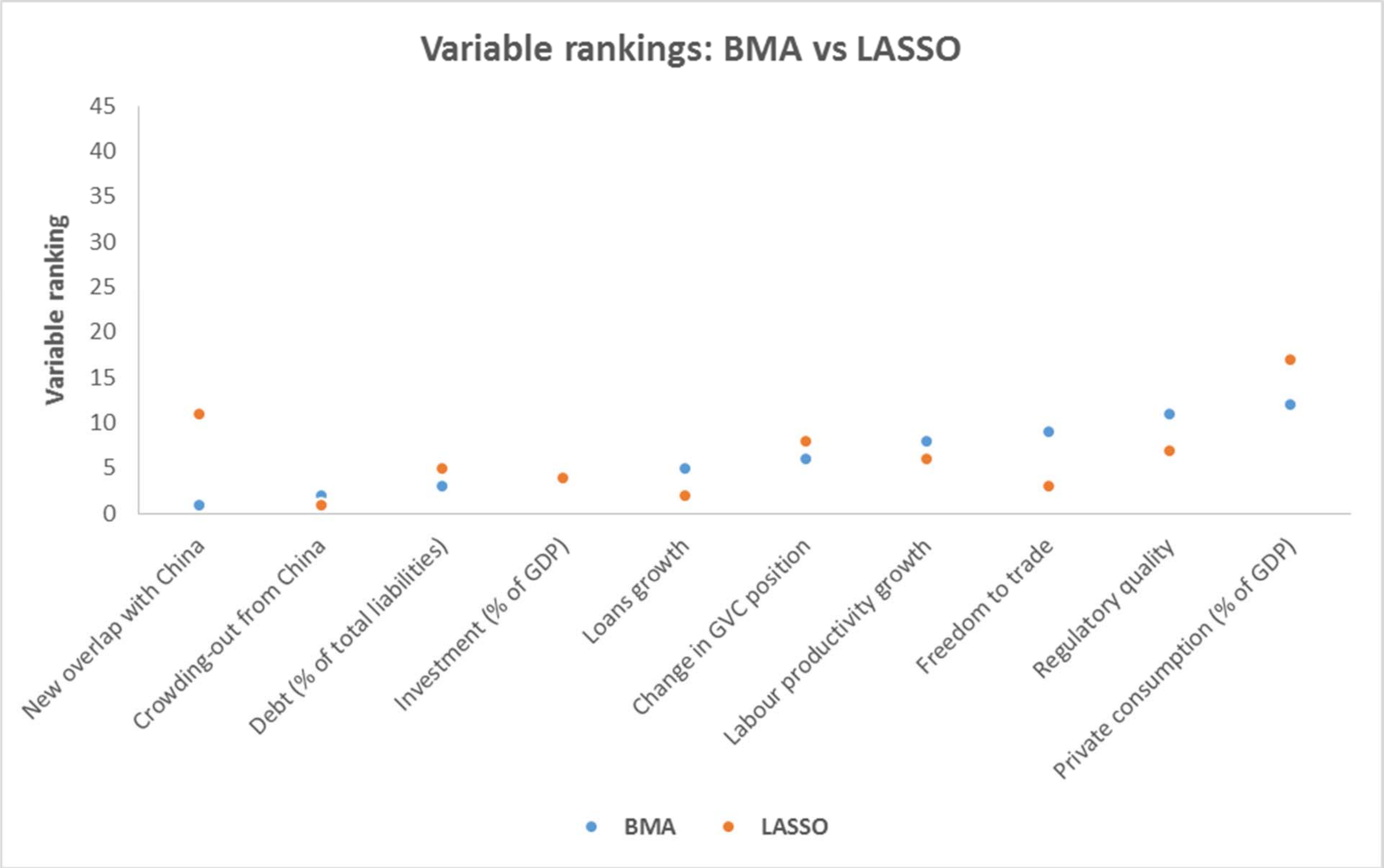
1. Start with all coefficients $\boldsymbol{\beta}$ equal to zero
2. Find the predictor x_j most correlated with y
3. Increase the coefficient β_j in the direction of the sign of correlation between x_j and y
4. Compute residuals $\varepsilon = y - \hat{y}$ along the way. Stop when some other predictor x_k has as much correlation with ε as x_j.
5. Repeat step 3. for (β_j, β_k), until some other predictor x_m has as much correlation with ε.
6. Continue until: all predictors are in the model

3.2. LASSO: Literature

- Theoretical foundations: Tibshirani, R. (1996), Efron, B., Johnstone, I., Hastie, T. and Tibshirani, R. (2003, Annals of Statistics)
- Applifations in forecasting: Bai and Ng (2008, JoE), De Mol, Giannone, Reichlin (2006), Buchmann (2011)
- Growth-literature: Schneider and Wagner (2008), Hal Varian (2014)

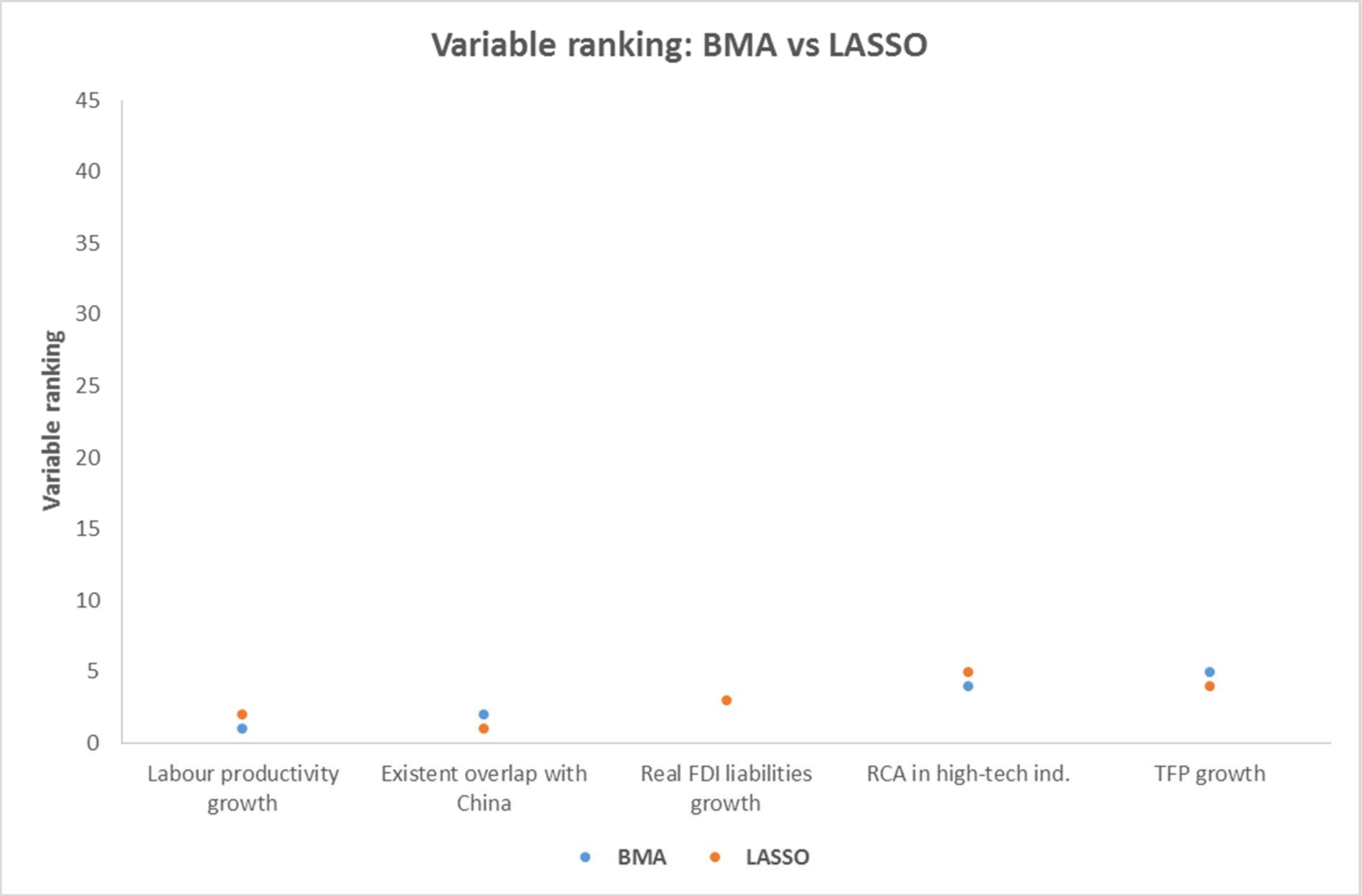
3.3. LASSO: Results

Old EU member states:



3.3. LASSO: Results

New EU member states:



3.3. LASSO: Results

Old EU member states	**BMA posterior mean**	**LASSO coefficient**
New overlap with China	-0,33	-0,32
Crowding-out from China	-0,36	-0,30
Debt (% of total liabilities)	-0,19	-0,18
Investment (% of GDP)	0,31	0,24
Loans growth	0,15	0,14
Change in GVC position	-0,16	-0,13
Existent overlap with China	-0,44	-
Labour productivity growth	0,16	0,09
Freedom to trade	0,20	0,16
TFP growth	0,22	0,12
Regulatory quality	0,13	0,10
Private consumption (% of GDP)	0,14	0,15
GVC participation	0,22	-
Public debt (% of GDP)	0,22	-
Legal system and property rights	0,11	0,11
Energy imports (% of energy use)	0,09	0,11
Size of government	-0,11	-0,13
Loans from foreign banks growth	-0,12	-0,06
Labour force with secondary ed.	0,10	0,10
HCI-CPI	-0,10	-0,06

3.3. LASSO: Results

New EU member states	BMA posterior mean	LASSO coefficient
Labour productivity growth	0,31	0,18
Existent overlap with China	-0,37	-0,21
Real FDI liabilities growth	0,22	0,09
RCA in high-tech ind.	-0,22	-0,07
TFP growth	0,28	0,05
Relative export prices adj. for quality	0,13	-
Labour force participation rate	-0,14	-
New overlap with China	0,18	-
Share of construction in investment	0,16	-
Tax burden	0,11	-
Loans from foreign banks growth	-0,14	-
Labour force with tertiary ed.	-0,13	-
Size of government	-0,11	-
GVC position	0,10	-
Patent applications	-0,10	-
Legal system and property rights	0,09	-
Growth surprise	0,04	-
Investment growth	0,09	-
Freedom to trade	0,07	-
Loans (% of total liabilities)	-0,10	-

Old EU member states: R-squared

BMA		LASSO	
AT	0,619	AT	0,852
BE	0,618	GR	0,740
GR	0,607	BE	0,699
DK	0,569	DE	0,640
IT	0,543	DK	0,563
DE	0,519	IT	0,534
NL	0,471	NL	0,522
FI	0,405	ES	0,463
ES	0,361	IE	0,405
SE	0,246	FI	0,344
IE	0,215	FR	0,292
PT	0,213	PT	0,290
FR	0,199	SE	0,275
UK	-0,074	UK	-0,040

3.3. LASSO: Results

New EU member states: R-squared

BMA		LASSO	
HU	0,657	HU	0,647
PL	0,640	PL	0,586
SI	0,568	SI	0,528
RO	0,422	RO	0,460
SK	0,421	CZ	0,421
BG	0,403	SK	0,373
CZ	0,398	BG	0,345
EE	0,388	LV	0,329
LV	0,349	EE	0,319
LT	0,247	LT	0,226
CY	0,238	CY	0,211

4. LASSO: Conclusions

Summary

1. LASSO yields very similar results when compared to BMA
   - For the new EU members states: Exact correspondence between indicators chosen by LASSO and robust indicators identified via BMA
   - For the old member states: From the 10 robust indicators according to BMA-methodology, LASSO ranks 8 among the top 10
   - LASSO coefficient signs are in line with BMA: Direction of effect does not seem to be driven by particular econometric method
2. R-squared suggests that exclusion of variables in combination with shrinkage (LASSO) can improve model fit
3. Next steps:
   1. Assess statistical significance of LASSO coefficients
   2. Conduct a country cluster analysis using the robust set of indicators

THANK YOU FOR YOUR ATTENTION!