Convergence Clubs and Regression Trees

0686 - Spatial Economics

Nikolas, Philipp, Lukas & Daniel

17 Januar, 2019

Data

Recap: European Regional Database by Cambridge Econometrics

We limit the dataset to:

- timeframe 2000-2015
- no Croatia (i.e. two fewer NUTS 2 regions)

This means we get to:

- use the full set of variables
- keep a detailled London (five NUTS 2 regions)

CS excursion

Regression vs. Iteration

Fibonacci sequence: $1, 1, 2, 3, 5, 8, 13, 21, ...$

Defined by

$$f(n) = f(n-1) + f(n-2)$$

with the base

$$f(1) = f(2) = 1$$

Regression tree

Split observations into clubs:

```
tree <- function(data, split_vars, end_criteria) {
  split <- find_best_split(...)
  if (!end_criteria) {
    return(list(tree(split$data1, ...),
                tree(split$data2, ...)))
  } else { # if(end_criteria)
    return(data)
  }
}
```

Oh what a merry regression tree

We receive a recursive, tree-like data structure that is:

- hard to deal with (**a lot** of helper functions are necessary)
- pretty nice

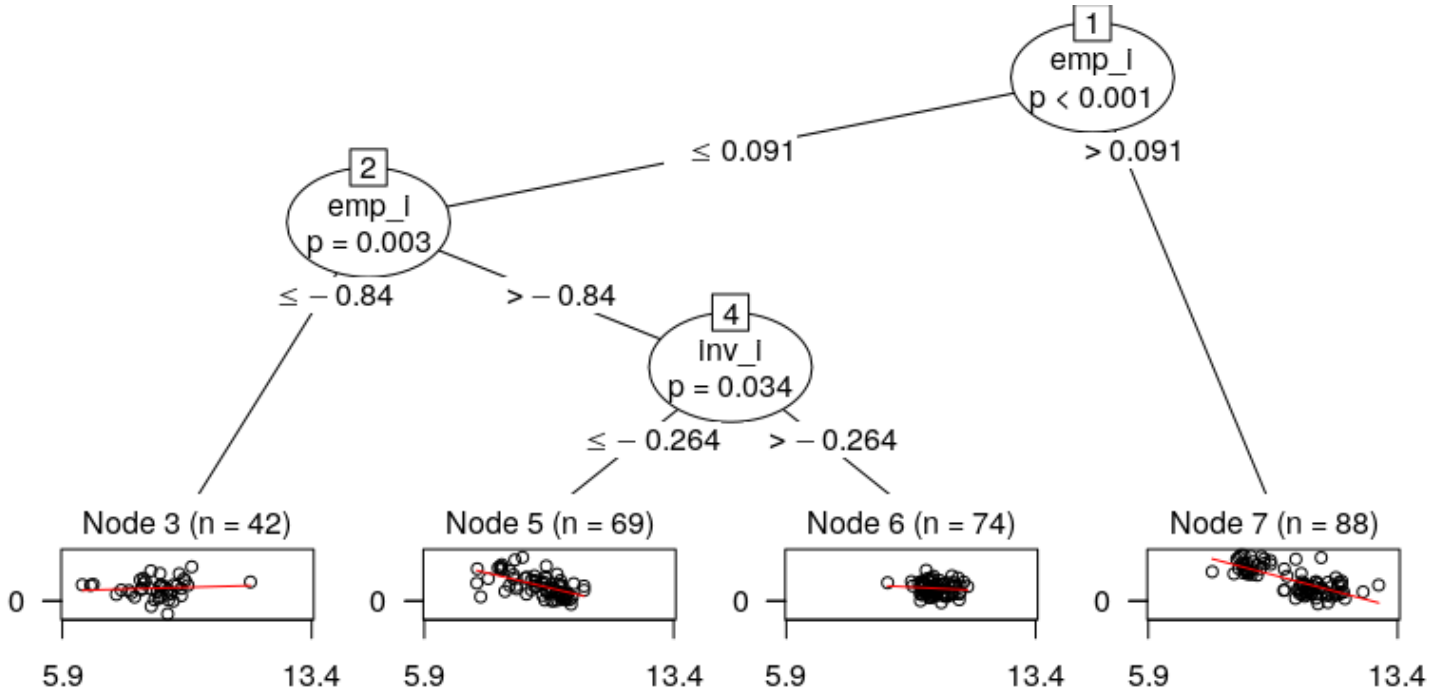


Figure 1: A partykit (Hothorn and Zeileis 2015) tree

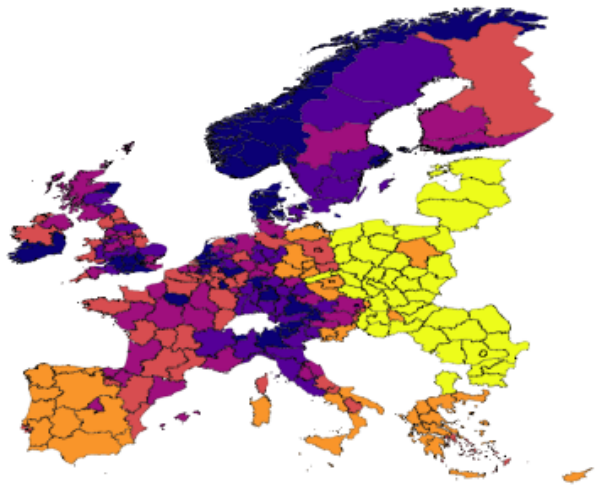
Regression Tree

Our results are comparable to partykit (Hothorn and Zeileis 2015).

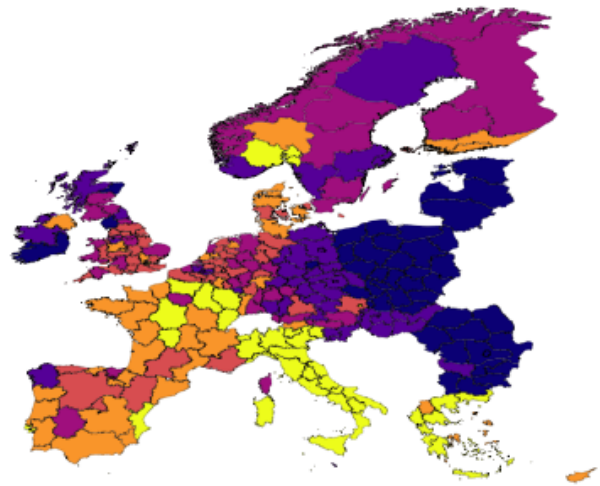
Still there's the caveat of spatially filtering the data.

Motivation

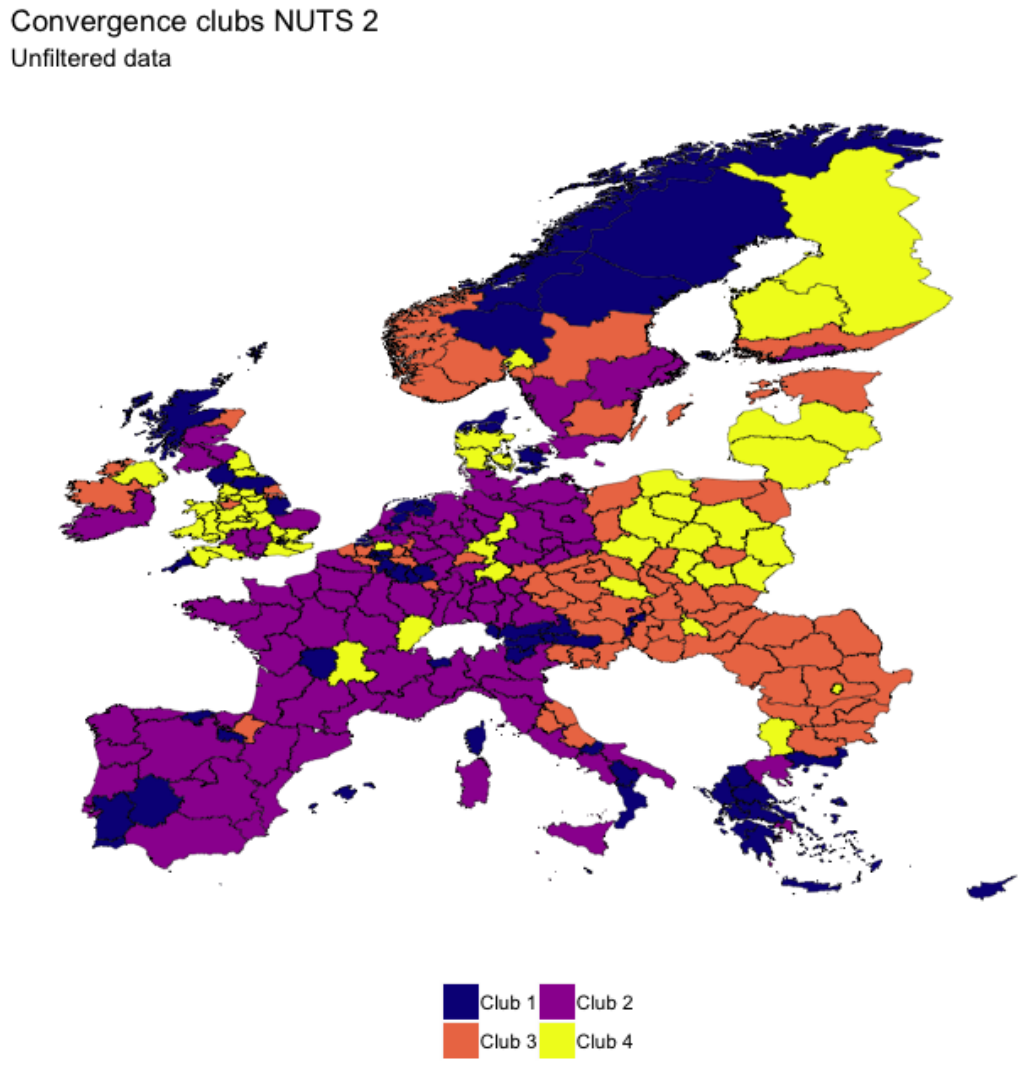
GDP p.c. in 2000
Quantile map



GDP p.c. growth 2000-15
Quantile map



Results


Convergence clubs NUTS 2
Unfiltered data
Club 1
Club 2
Club 3
Club 4

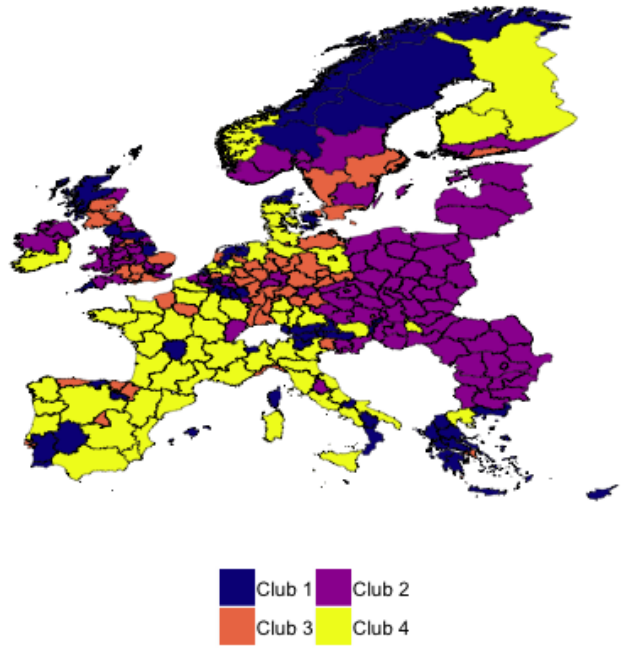
Results

Table 1: Regression results using unfiltered data

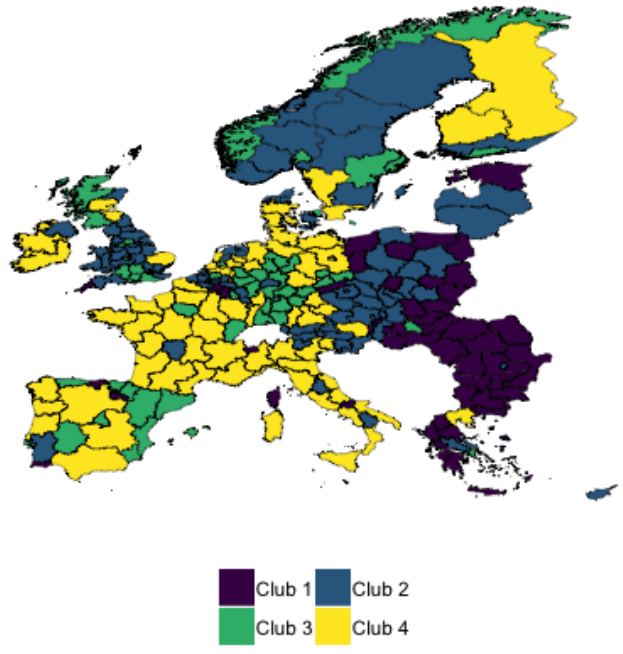
	Dependent variable:			
	GDP p.c. growth rate 2000-15			
	(1)	(2)	(3)	(4)
Constant	-1.139^{***}	-0.265	1.769^{***}	2.922^{***}
	(0.323)	(0.360)	(0.146)	(0.147)
Initial GDP p.c.	0.120^{***}	0.035	-0.159^{***}	-0.275^{***}
	(0.032)	(0.036)	(0.016)	(0.015)
Observations	63	92	67	51
Residual Std. Error	0.118 (df = 61)	0.105 (df = 90)	0.129 (df = 65)	0.086 (df = 49)
Note:				*p<0.1; **p<0.05; ***p<0.01

Results

Convergence clubs NUTS 2
SAR-filtered data



Convergence clubs NUTS 2
SEM-filtered data



Results

Table 2: Regression results using SAR-filtered data

	Dependent variable:			
	GDP p.c. growth rate 2000-15			
	(1)	(2)	(3)	(4)
Constant	-1.174^{***}	1.445^{***}	1.296^{***}	-0.037
	(0.343)	(0.122)	(0.383)	(0.470)
Initial GDP p.c.	0.109^{***}	-0.142^{***}	-0.128^{***}	-0.003
	(0.034)	(0.013)	(0.037)	(0.047)
Observations	63	97	55	58
Residual Std. Error	0.125 (df = 61)	0.124 (df = 95)	0.073 (df = 53)	0.110 (df = 56)
Note:				*p<0.1; **p<0.05; ***p<0.01

- partykit
- flattening trees
- try model on all?
- only spatial filtering

Results

Table 3: Regression results using SEM-filtered data

	Dependent variable:			
	GDP p.c. growth rate 2000-15			
	(1)	(2)	(3)	(4)
Constant	-0.039^{**}	0.088^{***}	0.016	-0.021
	(0.018)	(0.014)	(0.020)	(0.022)
Initial GDP p.c.	-0.277^{***}	-0.265^{***}	-0.061^{**}	-0.132^{***}
	(0.022)	(0.026)	(0.028)	(0.047)
Observations	55	89	59	70
Residual Std. Error	0.117 (df = 53)	0.120 (df = 87)	0.086 (df = 57)	0.106 (df = 68)

Note: *p<0.1; **p<0.05; ***p<0.01

- where are our clubs
- why are they split this way
- how do they compare (unfiltered, sar, sem)

Literatur

Hothorn, Torsten, and Achim Zeileis. 2015. “partykit: A Modular Toolkit for Recursive Partytioning in R.” *Journal of Machine Learning Research* 16: 3905–9.
http://jmlr.org/papers/v16/hothorn15a.html.