Convergence clubs and trees

Nikolas, Philipp, Lukas & Daniel

6 December 2018

Postiglione, Benedetti, and Lafratta (2010)
"A regression tree algorithm for the identification of convergence clubs"

Introduction

- Persistent country differences in GDP per capita[1]
- Hypothesis: convergence within clubs but not inbetween
- Concept of β-convergence:

$$ln(y_t/y_0) = \alpha + \beta ln(y_0) + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2 I)$$

- Spatial dependence likely[2]

[1] Barro and Sala-I-Martin 1995

[2] Anselin 1988

Regional convergence

- Consider spatial correlation or heterogeneity
- Filtered SAR-reformulation:

$$(I_n - \rho W) ln(y_t/y_0) = \alpha + \beta ln(y_0) + \omega_t, \quad \omega_t \sim N(0, \sigma_\omega^2 I)$$

- Filtered SEM-reformulation:

$$(I_n - \delta W) ln(y_t/y_0) = \alpha + \beta \left[(I - \rho W) ln(y_0)\right] + u_t, \quad u_t \sim N(0, \sigma_u^2 I)$$

Club identification

- Generally two different approaches:
  1. Select countries based on *a priori* criteria (e.g. GDP level)
  2. Consider the number of clubs as endogenous

A regression tree algorithm

- Non-parametric method to determine clubs
- Separate data using control variables
- Objective: find regions with similar coefficients

Regression trees

- Utilise binary recursive partitioning to subdivide a dataset
- Splits determined by imposed linear conditions on covariates
- In this case:
  - Find regions with similar parameters
  - Split if coefficients within clubs are significantly different

The algorithm

- Let Z be the splitting variables, with $z_{i,j}$ as variable i in region j

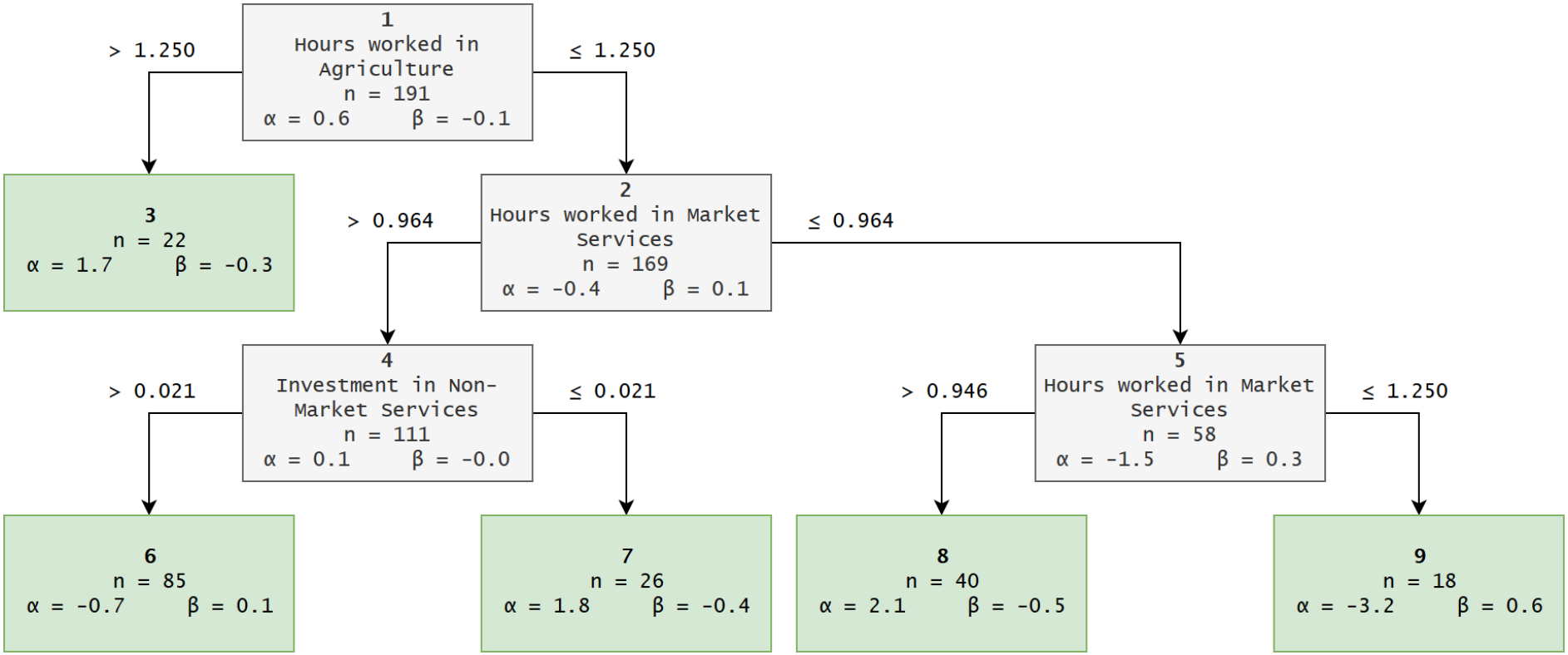
The algorithm

0. Start by estimating a pooled model for all EU regions

   At each step k, for each club C in the set of clubs S_k:

1. For every z_i partition the club s.t. $z_{i,j} > z_{i,curr}$ and $z_{i,j} \leq z_{i,curr}$
   - Re-estimate the model
   - Determine if the coefficients are significantly different
2. Use the best split (smallest p-value) for the next set of clubs S_{k+1}
3. End if:
   - No more significant differences
   - Minimum club size is reached
   - Maximum number of clubs is reached

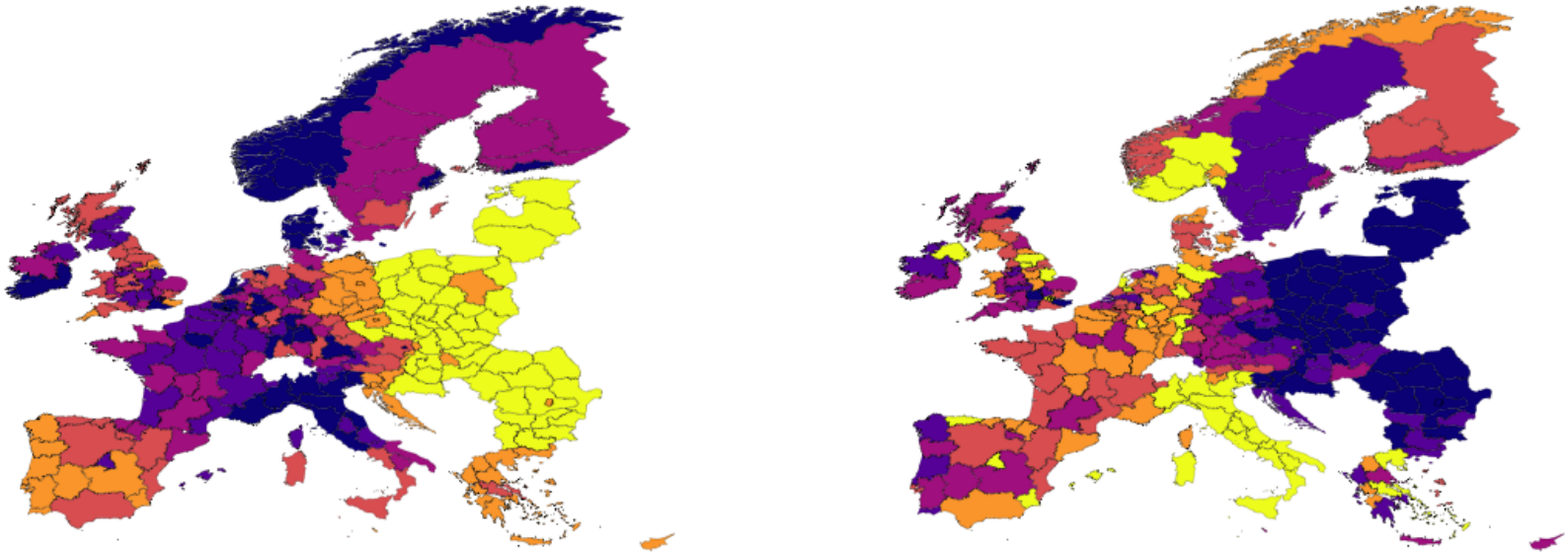
Tree results



Data

- Data obtained from European Regional Database by Cambridge Econometrics
- Timespan from 1980-2015; however, some regions only available from 1990 or 1995
- NUTS3 data only for certain indicators available - NUTS2 for all (NUTS2013)
- Available for 272 regions - yet to be determined if all will be included
- Data available for:
  - GVA, GDP
  - Employment indicators (total hours worked, compensation etc.)
  - Gross fixed capital formation
  - Population

Data Inspection



(a) Initial income levels

(b) Income growth rates

Figure: Quantile maps of initial level of income and subsequent growth rates

Spatial Dependence

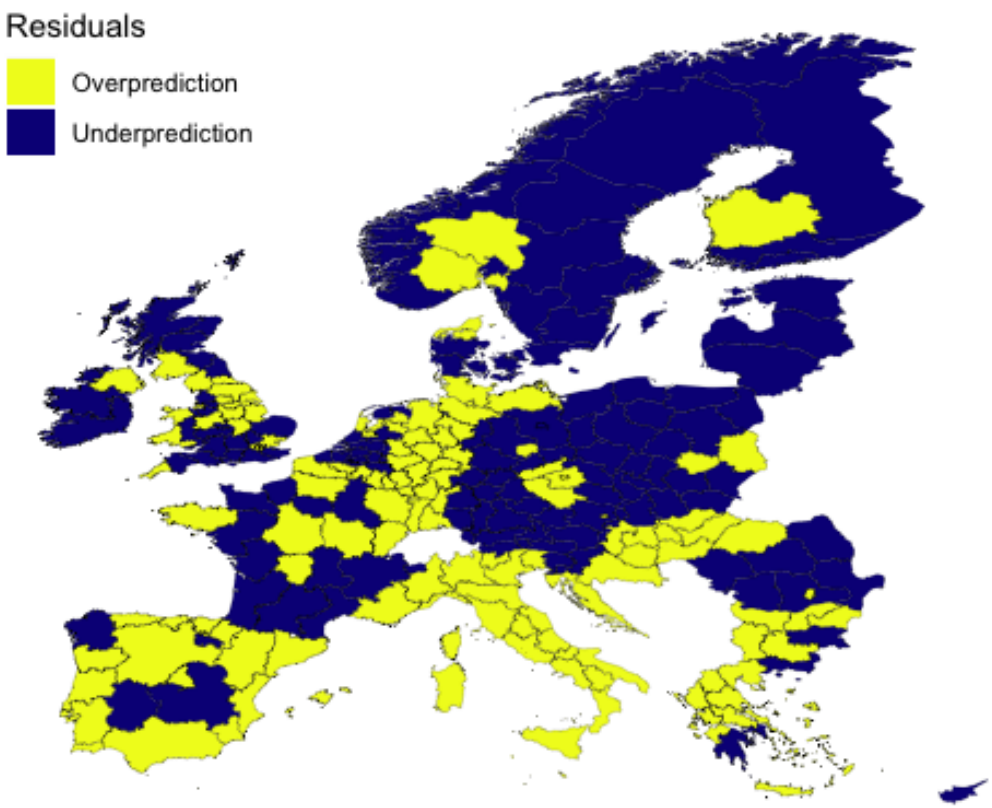


Figure: Map of OLS residuals with no club specification

References I

Anselin, Luc (1988). "Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity." In: *Geographical analysis* 20.1, pp. 1–17.

Barro, Robert J and Xavier Sala-I-Martin (1995). *Economic growth theory*.

Postiglione, Paolo, Roberto Benedetti, and Giovanni Lafratta (2010). "A regression tree algorithm for the identification of convergence clubs." In: *Computational Statistics & Data Analysis* 54.11, pp. 2776–2785.