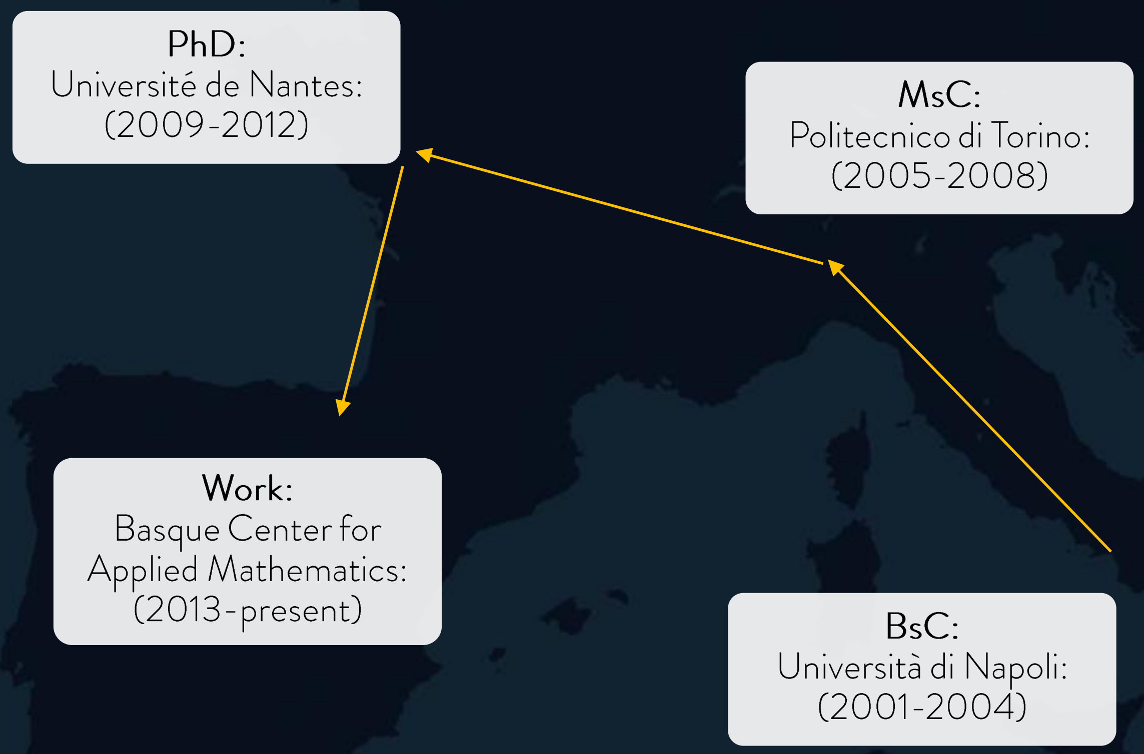
Researchers’ migration across Europe

What is guiding the flow of highly skilled workers?

Jacopo Ghirri, Carlo Ghiglione, Alfredo Gimenez Zapiola, Marta Mastropietro

ORCiD Data



Open Researcher and Contributor ID (ORCiD) associates with each researcher a numeric code.

Keeps track of all publications, academic and working career.

1. Migration trajectories of all enrolled researchers;
2. Academic migration flows across 35 European countries during 2006-2015.

Data for each country

Migration flows by year:

- Inflow
- Outflow

University score:

- Ad-Hoc designed indicator of the overall quality of universities

Education-specific:

- Expense in R&D over GDP
- Teachers for student ratio
- …

Socio-economic:

- GDP per capita
- Human Development Index
- …

University Score

COUNTRY: *ITALY*

(QS World University Rankings 2018)

- class 1 [1, 100]: 0
- class 2 [101, 200]: 4
- class 3 [201, 300]: 2
- ...
- class 9 [801, 900]: 0
- class 10 [901, 1000]: 5

UniScore = 0.248

University Score is a **weighted sum** of the number of universities in specific positions of the **QS World University Ranking**.

$$UniScore = \sum_{c=1}^{10} \frac{1}{c} n(c)$$

with $n(c)$ the **number of universities** in the c^{th} class for the country.

It is an indicator of the **overall academic prestige** of a country.

Goal of the analysis

The research question can be declined into:

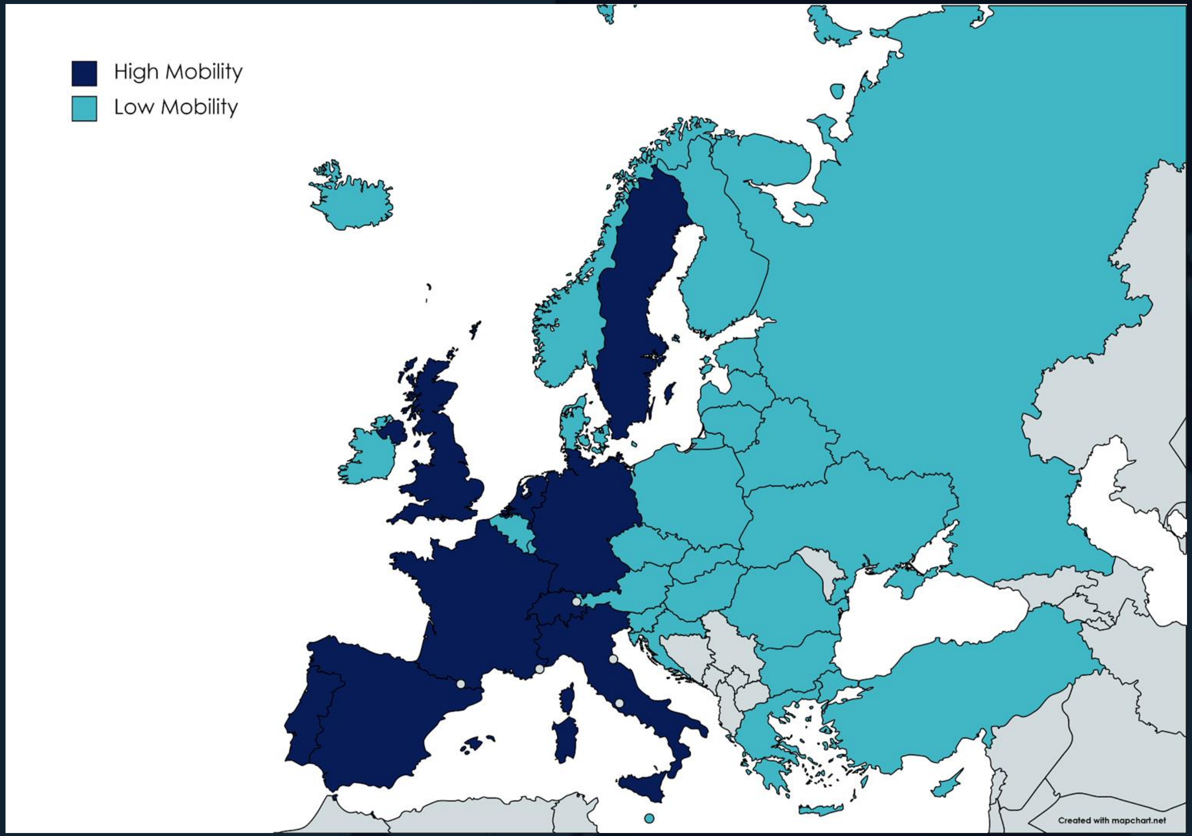
1. Determine the interactions and differences, if any, between groups of countries;

2. Highlight the principal determinants that influence flows in and out of a country, finding appropriate models;

3. Understand whether there is a temporal characterization or evolution.

Analysis of determinants by groups

Clustering divides Europe in **high mobility** and **low mobility** countries.



Problem: homoschedasticity assumption of ANOVA is **strongly violated**.

⇩

Solution: Nonparametric Combinations based ANOVA

⇩

Multi-aspect testing applying a p-value correction for the two tests via Nonparametric Combination.

Analysis of determinants by groups

We find a prototype of the general high mobility country.

Its features are:

1. Good universities
2. Good quality of life
3. Active entrepreneurial research
4. Investments

Surprisingly, no significant differences in terms of general wealth of a country (GDP per capita) or the general level of education of its population.

Regressor	p-value
University Score	0.000
Citations per document	0.100
University staff	0.011
Students per staff	0.945
GDP per capita	0.097
R&D over GDP	0.048
Education index	0.759
Local Purchasing Power Index	0.045
Global Gender Gap Index	0.042
Public Services	0.031
Security Apparatus	0.201
Patent Application	0.015

Determinants Analysis

Objective: find the determinants of the **flows** of researchers.

⇩

First stage: covariate selection with both linear regression and GAMs.

⇩

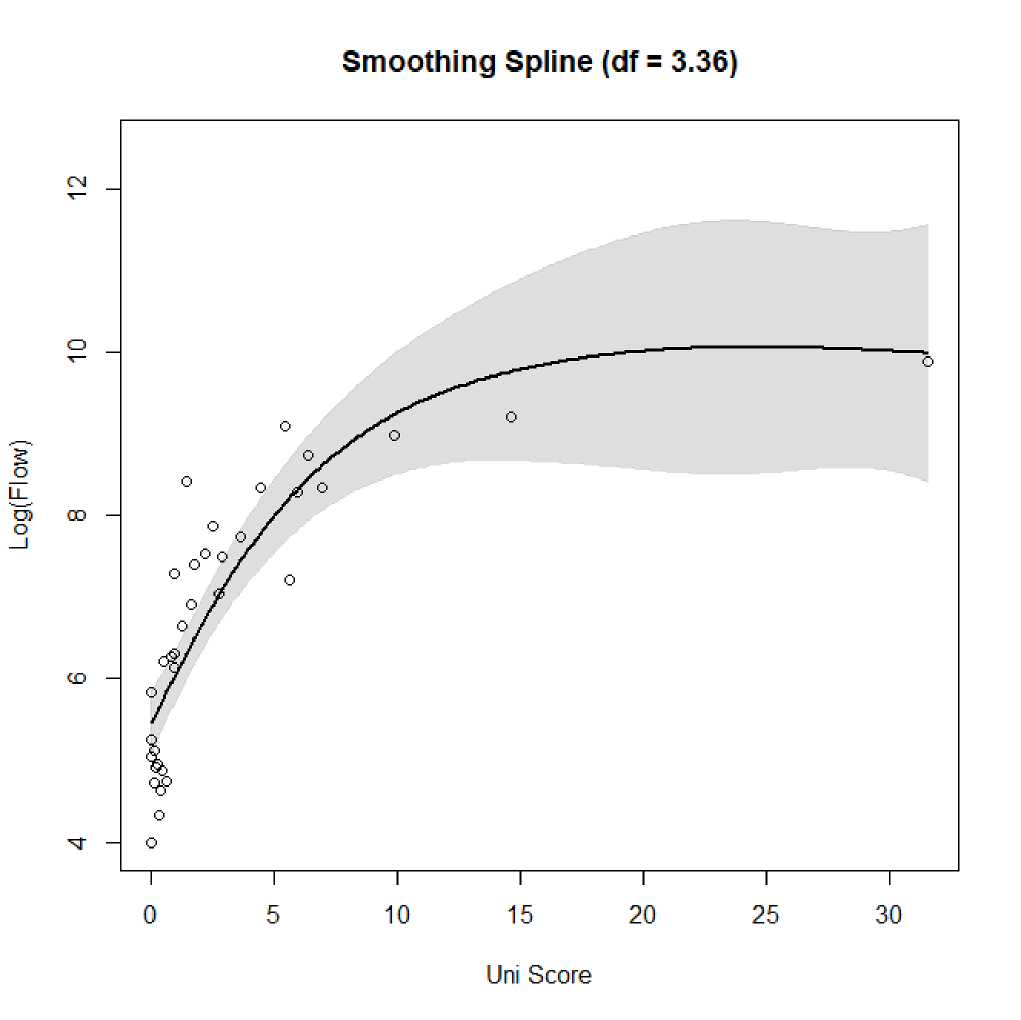
Second stage: tuning of parameters with AIC and stratified k-fold cross validation selection criteria.

⇩

Multiple nonparametric regression models.

Selected Models

Model for total flow

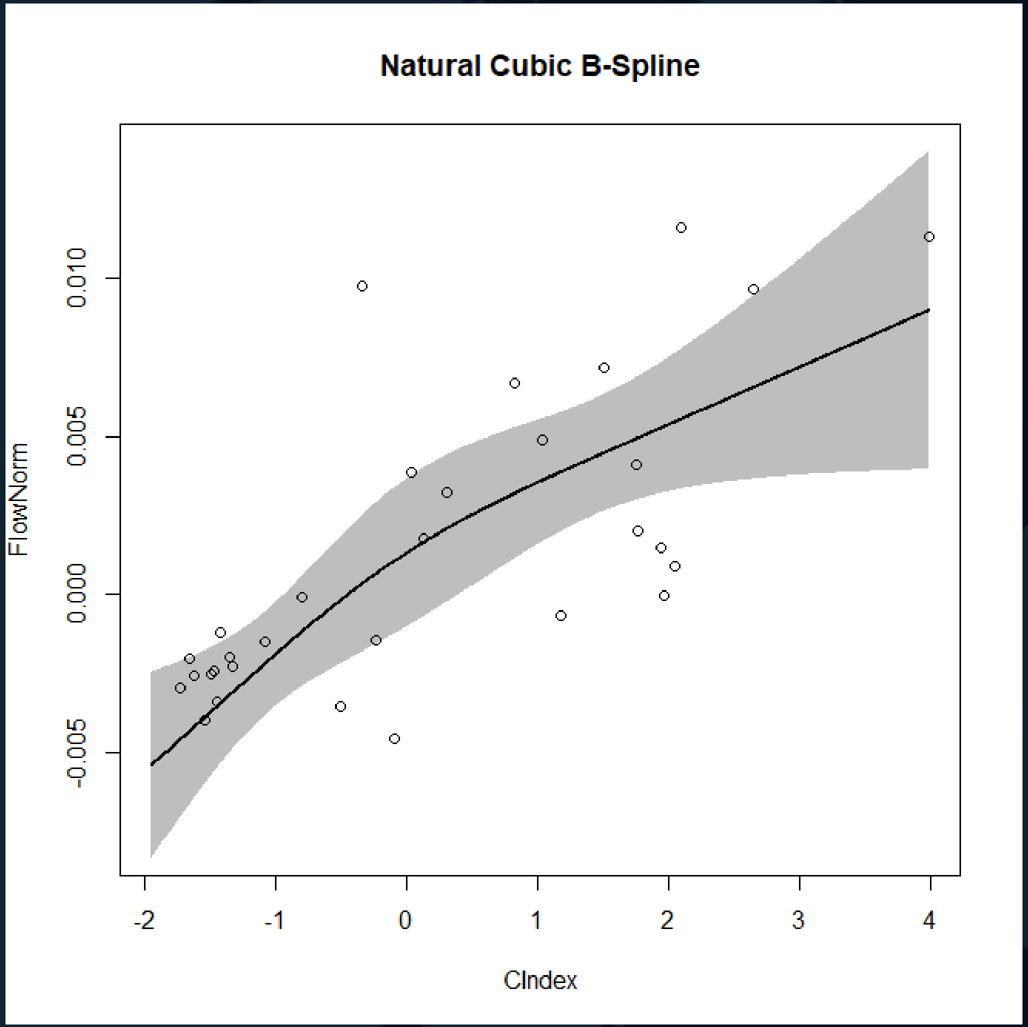


University score alone explains around 75% of the variability in total flow (using logarithmic scale).

The prestige of a country university is a fundamental driver for low mobility countries but reaches a plateau for the already active ones.

Selected Models

Model for Normalized Flow



Net flows (in - out) are normalized wrt the number of researchers of the country.

PCA built Composite Indicator of:

- Citations per Document
- Academic staff per student
- %GDP in R&D
- University Score

Almost linear dependence emerges even when using more sophisticated models.

Temporal Analysis – Repeated Measures

Objective: determining whether the phenomenon changes over time.

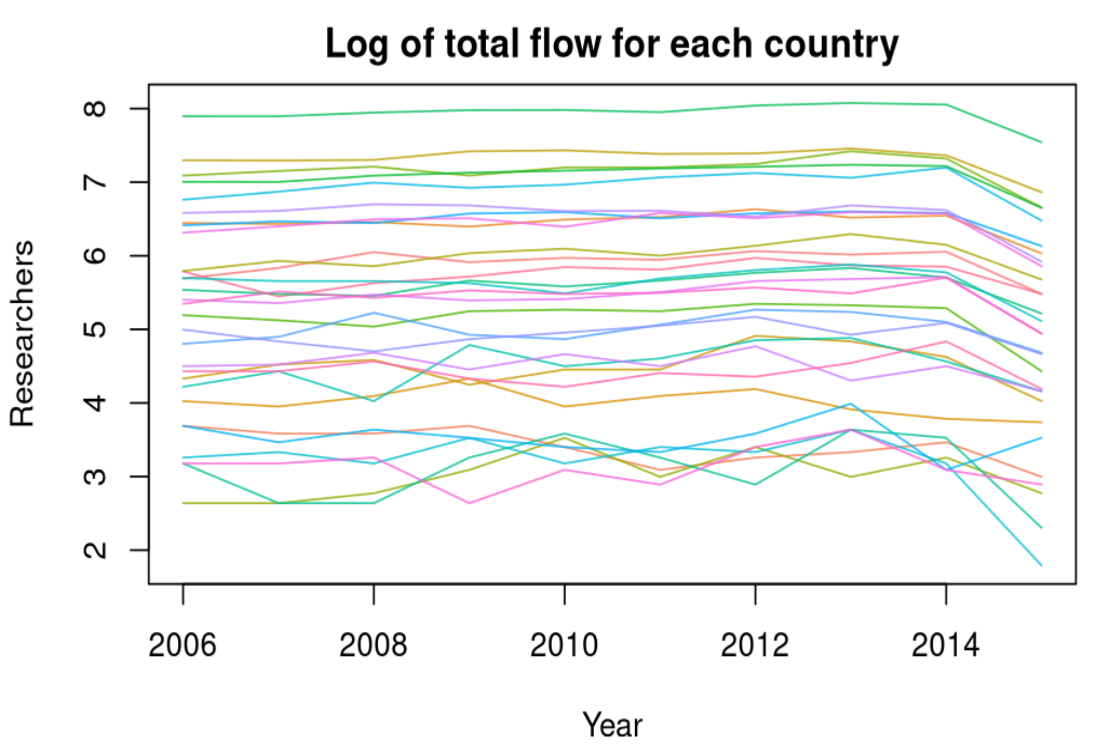
⇩

Approach: repeated measures for each country along time.

⇩

Robust Mixed effects model:

interaction of *UniScore* and time is tested, treating dependence between observations of the same country with a random term.



Temporal Analysis – Repeated Measures

We use the following model:

$$\boldsymbol{y}_i = X_i \boldsymbol{\beta} + Z_i \boldsymbol{b}_i + \boldsymbol{\varepsilon}_i \quad \forall i \in Country$$

$$\boldsymbol{b}_i \overset{ind}{\sim} N\left(\boldsymbol{0}, Z_i D_i Z_i^T\right) \quad \forall i \in Country$$

$$\boldsymbol{\varepsilon}_i \overset{iid}{\sim} F_\varepsilon$$

Robust approach: residuals are weighted with a smoothed Huber function (95% efficiency), same for their scale.

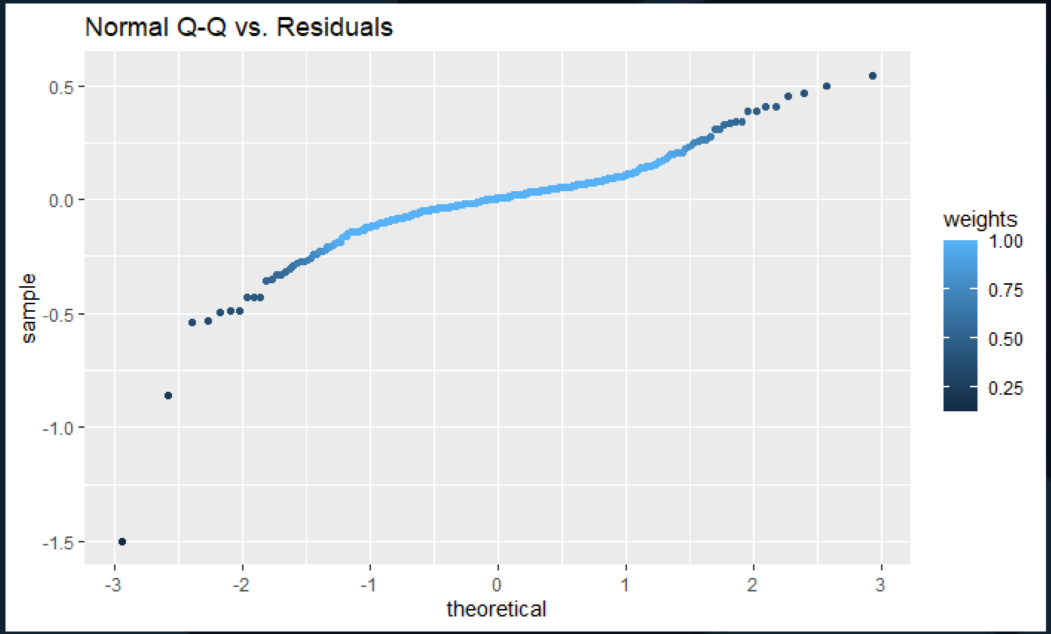
Permutation test:

$$H_0 : \log\left(TotFlow_{i,t}\right) = \beta_0 + f(UniScore_i) + \tilde{\varepsilon}_{i,t} \quad \forall t \in Year, \forall i \in Country$$

$$H_1 : \log\left(TotFlow_{i,t}\right) = \beta_0 + f_{\mathrm{t}}(UniScore_i) + \tilde{\varepsilon}_{i,t} \quad \forall t \in Year, \forall i \in Country$$

$f(x)$ *linear spline*

Temporal Analysis – Repeated Measures



QQ plot of the residuals with respective weights given by the smoothed Huber function.

P-value = 0

We reject the null hypothesis.

The effect of university score changes over time.

Temporal Analysis - FDA

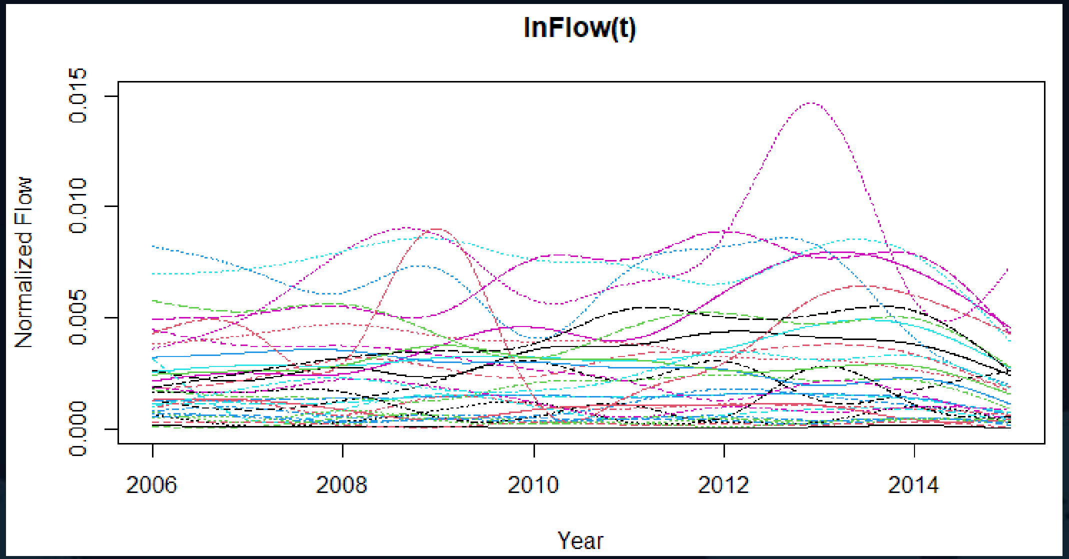
Objective: Study the **concordance** of in & out flows of researchers across the years.

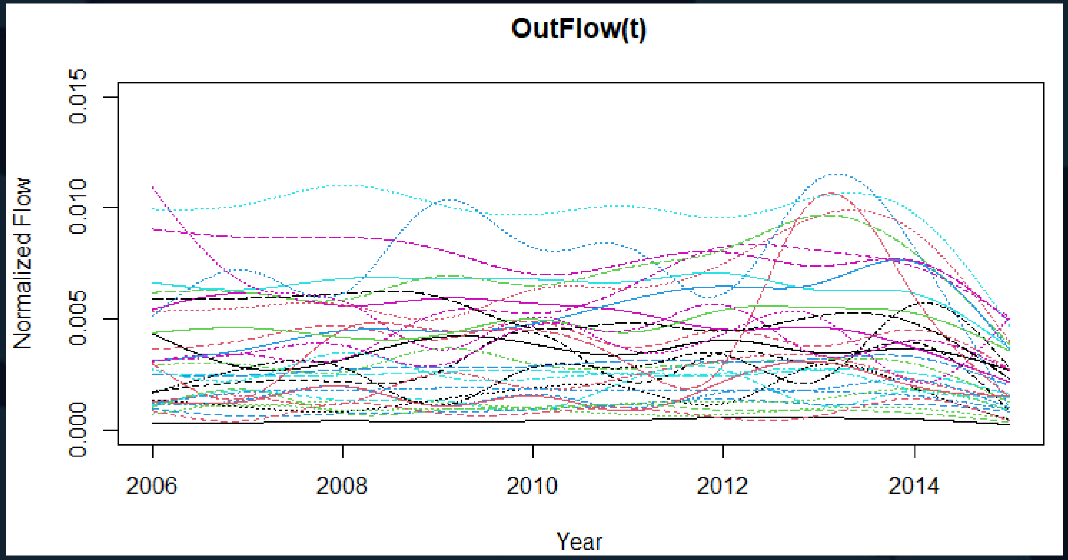
⇩

Approach: Functional Data Analysis.

⇩

Normalized flows wrt the number of researches in the country (to address **scale variability** issues).





Temporal Analysis - FDA

Independence test of two functional populations:

- $X(t)$ normalized inflow
- $Y(t)$ normalized outflow

Permutation test on absolute value of Spearman Correlation index $\boldsymbol{\rho}(X(t), Y(t))$

$$H_0: \ | \boldsymbol{\rho}(X(t), Y(t)) | = 0$$

$$H_1: \ | \boldsymbol{\rho}(X(t), Y(t)) | \neq 0$$

P-value = 0

⇩

Inflow and outflow are not uncorrelated.

$\boldsymbol{\rho}(X(t), Y(t)) = 0.942$

⇩

They behave in a positive concordant way.

Policy proposals

Low Mobility Countries

- Lack of prestige of universities is detrimental for the researchers mobility.
- Already existing EU funds for R&D should be targeted at improving local universities.

High Mobility Countries

- Further improvements of already prestigious institutions are not that efficient by themselves.
- Generalized investments on activity and personnel expansions can increment the attractivity.

Bibliography

- Caughey, Devin, Dafoe, and Seawright. Nonparametric Combination (NPC): A framework for testing elaborate theories. *The Journal of Politics*, 79(2):688–701, April 2017
- N. M. Laird and J. H. Ware. Random-Effects Models for Longitudinal Data. *Biometrics*, 38(4):936–974, December 1982.
- J. C. Pinheiro and James H. Ware. Conditional versus Marginal Covariance representation for Linear and Nonlinear models. *Austrian Journal of Statistics*,35(1):31–44, April 2016.
- F. Ieva, F. Palma, and J. Romo. Bootstrap-based Inference for Dependence in Multivariate Functional Data. *Submitted*, 2017
- M. Koller. robustlmm: An R Package for Robust Estimation of Linear Mixed-Effects Models. *Journal of Statistical Software*, 75(6):1–24, December 2016.