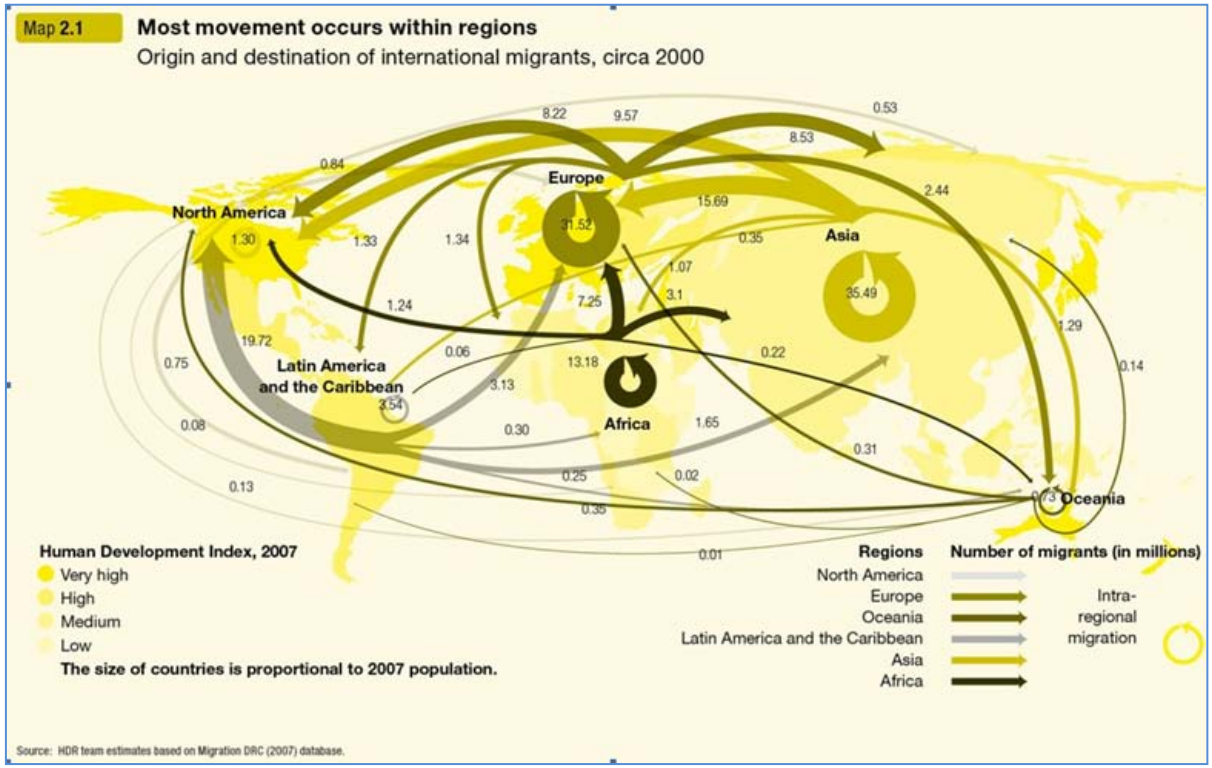
CUCS
MILANO 2017
V CONGRESSO

A model of the role of education in 2015 UN international migration data

Giorgio Guariso, Giacomo Toffano
DEIB, Politecnico di Milano

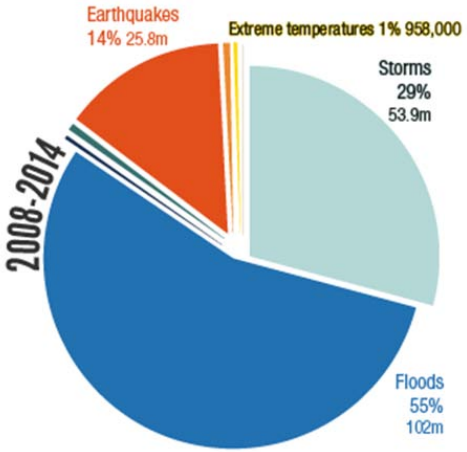
International migration is an important (and evolving) problem

From 2.8% of the world population in 2000 to 3.3% in 2015 (244 million international migrants). A higher growing rate than world population.



+ 740 million of migrants within states.

183 million (Int. Displacement Monitoring Center) 2008-14 for climatic reasons.



The decision to migrate can be examined under many perspectives

Psychology, Economics, Geography, Politics, History, Ethnology, Demography have contributed to the analysis.

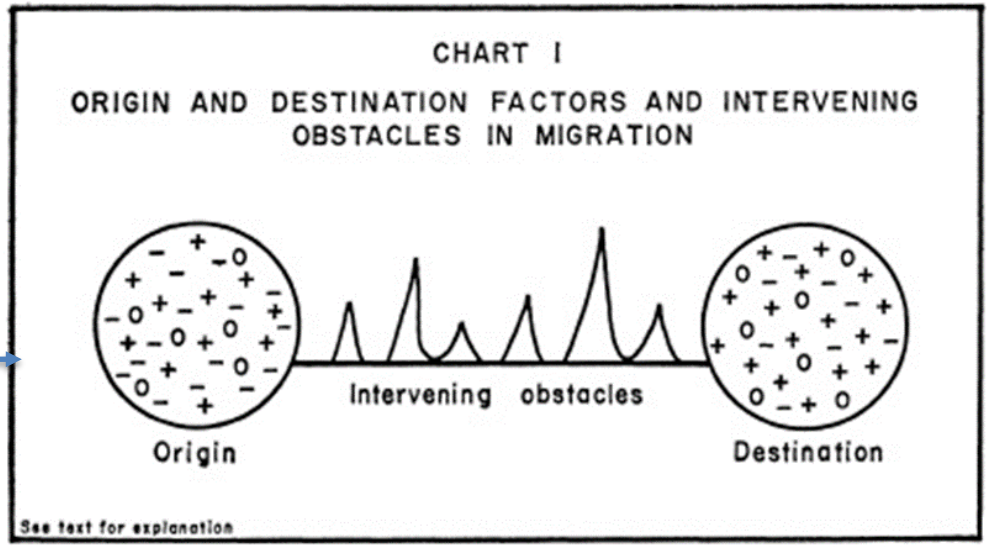
Two approaches

- **Macro** level: movement patterns of the population of a nation/area based on **census data**
- **Micro** level: formalization of the **decision process** of individual migrants (agent-based model)

Ravenstein 1885: «Laws of Migration»

Lee 1966: Push-Pull Migration Model

Push factors (origin), pull factors (destination), obstacles



The extended gravitational model

Following Lee's perspective, one may try to relate some indicator of migration intensity M to push and pull factors P in the area of origin i and destination j using a formula similar to the Newton's gravitational law

Push and pull factors

Migration indicator

$$M_{ij} = G \cdot \frac{P_i . P_j}{D_{ij}^2}$$

constant

Distance between the two areas

This can be generalized to

$$M_{ij} = G \cdot \frac{P_i^{\beta 1} . P_j^{\beta 2}}{D_{ij}^{\alpha}}$$

where α and β parameters have to be estimated to best fit actual data.
More push and pull factors can also be written in the numerator.

The tested model (log transformed to be linear)

Dependent variable :

- $F_{ij} = \frac{M_{ij}}{P_i}$ → M_{ij2015}
  → $M_{ij\Delta} = M_{ij2015} - M_{ij2000}$

Regressors

Component of the Human Development Index (HDI)

- $LifeExp$
- $SchoolExp$
- $SchoolMean$
- GDP
- EMP

Economic variables

- D_{ij}
- $border$

Geographic variables

- $INEQ$
- AGE

Equation:

$$\log F_{ij} = \log \alpha + \beta_0 \log GDP_i + \beta_1 \log GDP_j + \beta_2 \log D_{ij} + \beta_3 \log AGE_i + \beta_4 \log AGE_j$$

$$+ \beta_5 \log LifeExp_i + \beta_6 \log LifeExp_j + \beta_7 \log SchoolExp_i + \beta_8 \log SchoolExp_j$$

$$+ \beta_9 \log SchoolMean_i + \beta_{10} \log SchoolMean_j + \beta_{11} \log INEQ_i + \beta_{12} \log INEQ_j + \beta_{13} \log EMP_i$$

$$+ \beta_{14} \log EMP_j + \beta_{14} border$$

Data: UN statistical subdivision

The world is subdivided in 18 macro areas (Caribbean, Micronesia, Melanesia, and Polynesia were disregarded) → a matrix of 18 x 18 = 324 flows

Method and results

Too many "independent" variables correlated to each other → **pruning**

A repetitive procedure in which, at each step:

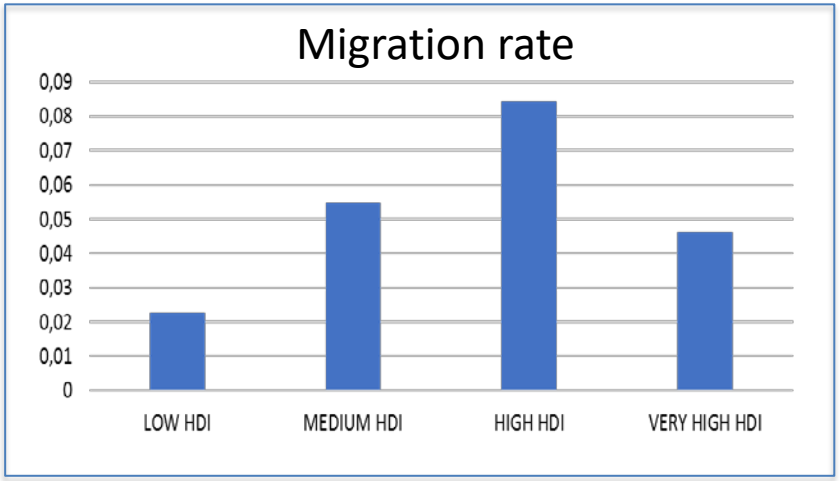
- the least significant parameter (and related variable) is eliminated
- all the other parameters are calibrated anew
- the new parameters are checked for statistical significance

In the end, only four variables always remain:

	Resident migrants		Variation resident migrants	
Correlation R data-model	0,758		0,734	
R^2	0,574		0,538	
	Model coefficients			
	Value	*Significance*	*Value*	*Significance*
Distance (log)	-1,787	6,27E-10	-1,792	7,94E-09
Border	0,67	9,93E-05	0,505	6,27E-03
GDP destination (log)	2,007	4,14E-41	1,951	2,45E-33
exp school origin (log)	3,95	6,81E-12	2,199	2,68E-04

Conclusions: relevant variables

- The same variables remain in all the regressions → robustness
- Distance and border: geographical structures (negative and positive coefficients)
- Expected GDP destination at destination: economic pull (positive coefficient)
- **Expected degree of education in the area of origin** (migrant selectivity) (highly positive)



There is a known NONLINEAR relation between some of the considered variables (e.g. education, GDP origin) and migration rates that calls for different modelling approaches.