INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

» Labour Market Attractiveness in the EU

HACKATHON FOLLOW-UP 2017 «

Sonia Quaresma, Joao Lopes and Marco Moura

(18-09-2017)

»

Motivation

- "Skills development are essential in the emerging new economy and fast-changing labour market"[1]
- "Qualification and skill mismatches entail significant economic and social costs for individuals and firms"[1]
- Skills mismatch (i.e. over-qualification, under-qualification) remains at 45% (CEDEFOP, 2015)[2]
- EU Guidelines (2015) call for enhancing labour supply, skills and competences[3]

Motivation

Create framework that:

1. combines **Official Statistics** with **Big Data**

2. estimates **Labour Market Attractiveness** and its association with **Skills Mismatch**, **Labour Market Mobility** and **Emigration**

3. is aimed at **policy makers** and both **jobseekers** and **job providers**

Data: LMkt Attractiveness

- “reg_dem” – demographic statistics
- “earn” – earning structure
- “educ_uoe_fin” – public expenditure on education
- “ilc” – income and life conditions
- “employ” – employment statistics
- “nama10” – annual national accounts
- “educ_part” – participation in education

7 datasets, **17** main variables

Data: LMkt Attractiveness

- "reg_dem" by **age** (**NUTS2**)
- "earn" by **occupation** and **economic activity**
- "educ_uoe_fin"
- "ilc" (**NUTS2**)
- "employ" by **age**, **education level**, **economic activity** (**NUTS2**)
- "nama10" (**NUTS2**)
- "educ_part" (**NUTS2**)

7 datasets, **17** main variables, **76** variables

subjects: **NUTS0 = 28**; NUTS1= 98; NUTS2 = 276.

Data: Skills mismatch

- EURES **scrapped data** on jobseekers' CVs
- EURES **scrapped data** on job vacations

- ESCO to **map** qualifications and occupations/economic sectors

2 datasets, **1** main variable

... but

cleaning and **structuring** requires considerable expertise

normalization requires detailed demographical information

mapping not available yet

Data: Skills mismatch

- “educ_uoe_grad02” – education statistics
- “jvs_a_nace2” – job vacancy statistics

- *ad hoc* mapping

2 datasets, **1** main variable

Data: Skills mismatch

- "educ_uoe_grad02" by **education field**
- "jvs_a_nace2" by **occupation** and **economic activity** (**NUTS2**)

- *ad hoc* mapping

2 datasets, **1** main variable, **1** variable

subjects: **NUTS0 = 8**; NUTS1 = 14; NUTS2 = 47

Data: LMkt Mobility

- "lfso_14leeow" – labour force statistics

1 datasets, **1** main variable

Data: LMkt Mobility

- "lfso_14leeow" by **occupation** and **education level**

1 datasets, **1 main variable**, 14 variables

subjects: **NUTS0 = 25**

Data: Emigration

- “migr_emi2” – migration statistics

1 datasets, **1** main variable

Data: Emigration

- "migr_emi2" by **age**

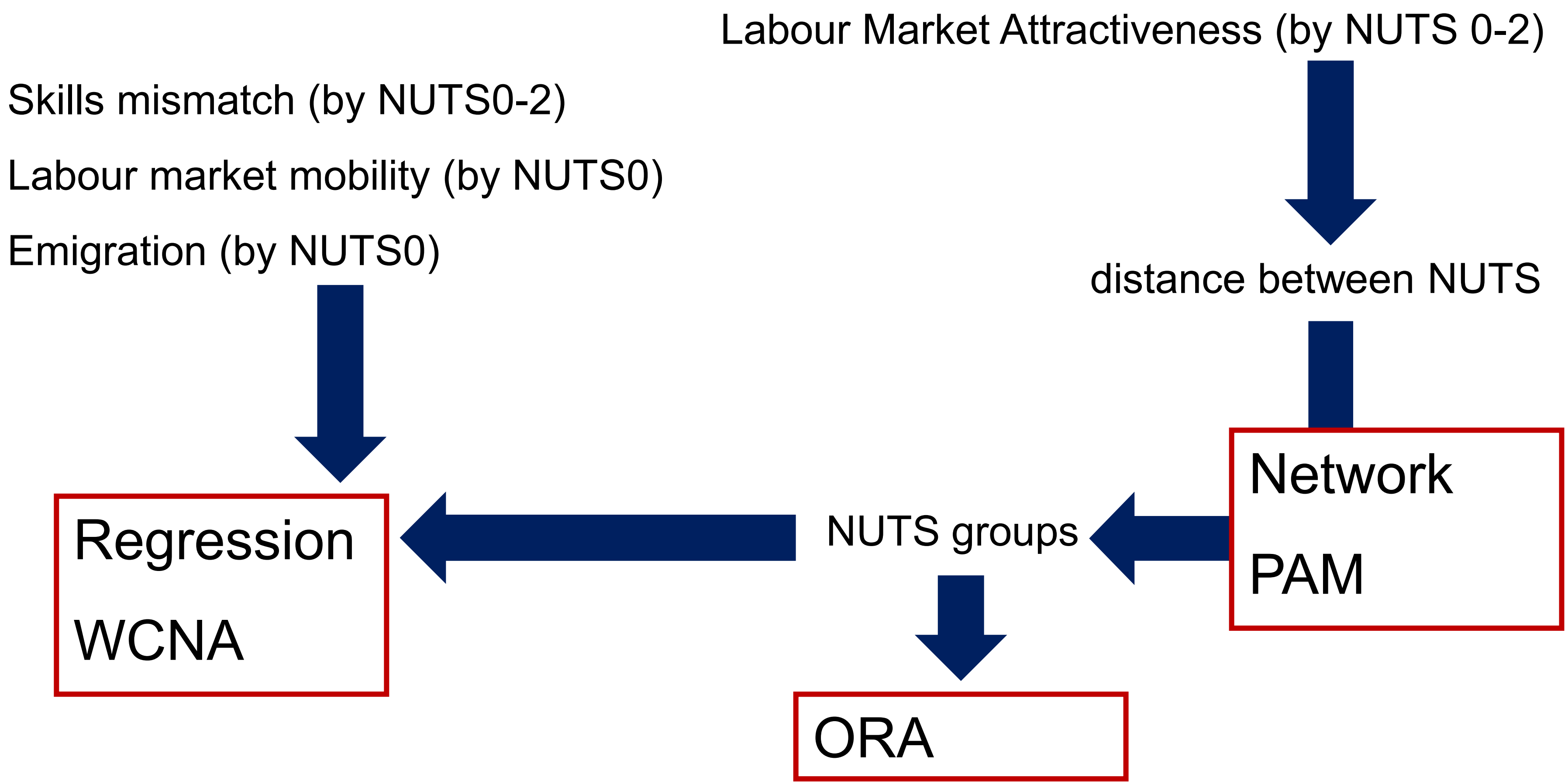
1 datasets, **1 main variable**, 3 variables

subjects: **NUTS0 = 28**

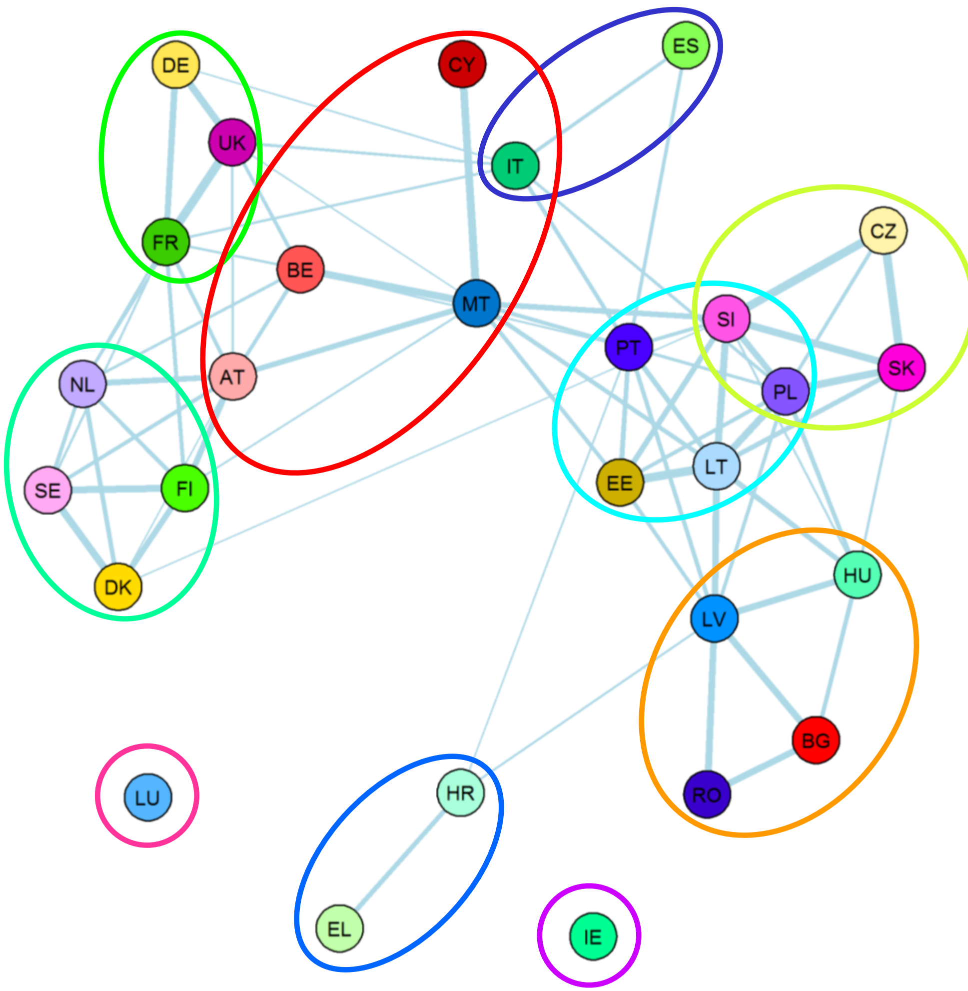
Methods

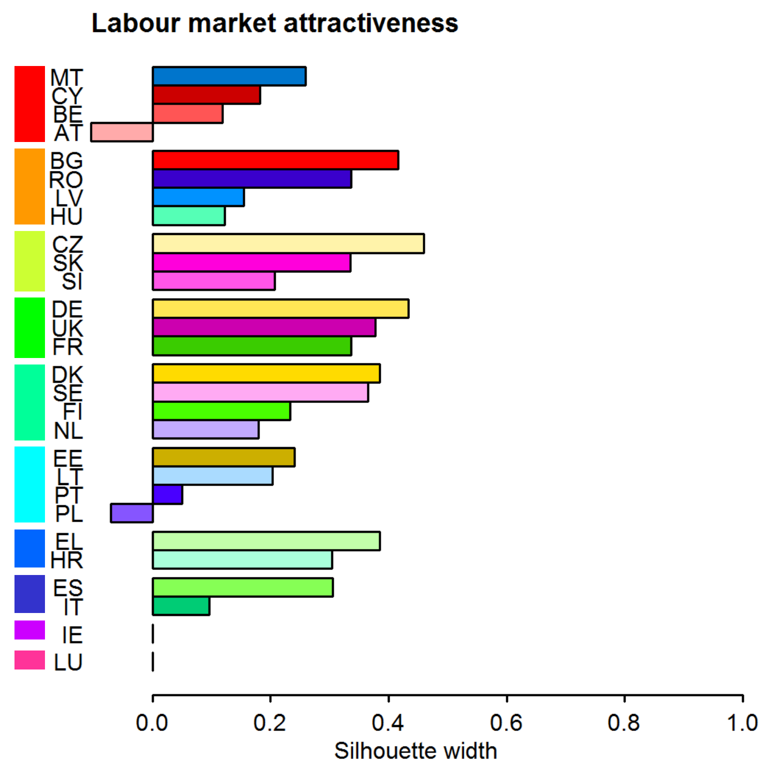
- Social network analysis
- Partition-around-medoids (PAM)[4]
- Over-representation analysis (ORA)
- Multinomial logistic regression
- Multivariate linear regression
- Weighted correlation network analysis (WCNA)[5]

Methods


Labour Market Attractiveness (by NUTS 0-2)
Skills mismatch (by NUTS0-2)
Labour market mobility (by NUTS0)
Emigration (by NUTS0)
distance between NUTS
Network
PAM
NUTS groups
Regression
WCNA
ORA

Results: LMkt Attractiveness





Results: LMkt Attractiveness

"MT-CY-BE-AT"
emp_15-24_ED5-8
pop_Y15-24

"BG-RO-LV-HU"
ARPR_socexcl
disp_income
expend_ED5-8

"CZ-SK-SI"
emp_Y25-64_ED3-4
pop_Y25-64
emp_Y25-64_ED0-2

"DE-UK-FR"
disp_income
pop_Total
pop_Y0-14

"DK-SE-FI-NL"
emp_Y[15-24,25-64]

expend_ED5-8
ARPR_socexcl

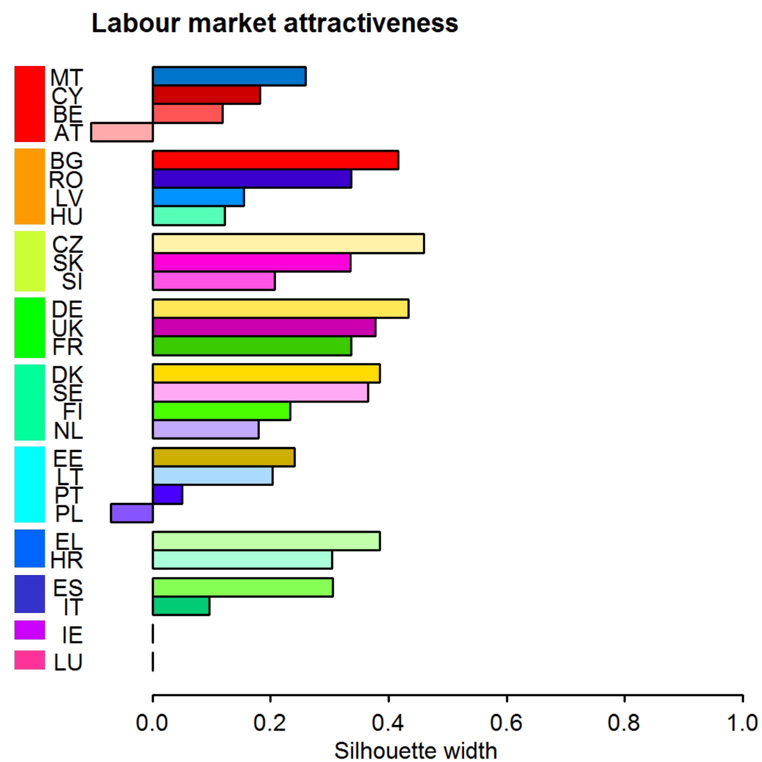
"EE-IT-PT-PL"
emp_YGE65

"EL-HR"
unemp_Y15-24
emp_Y25-64

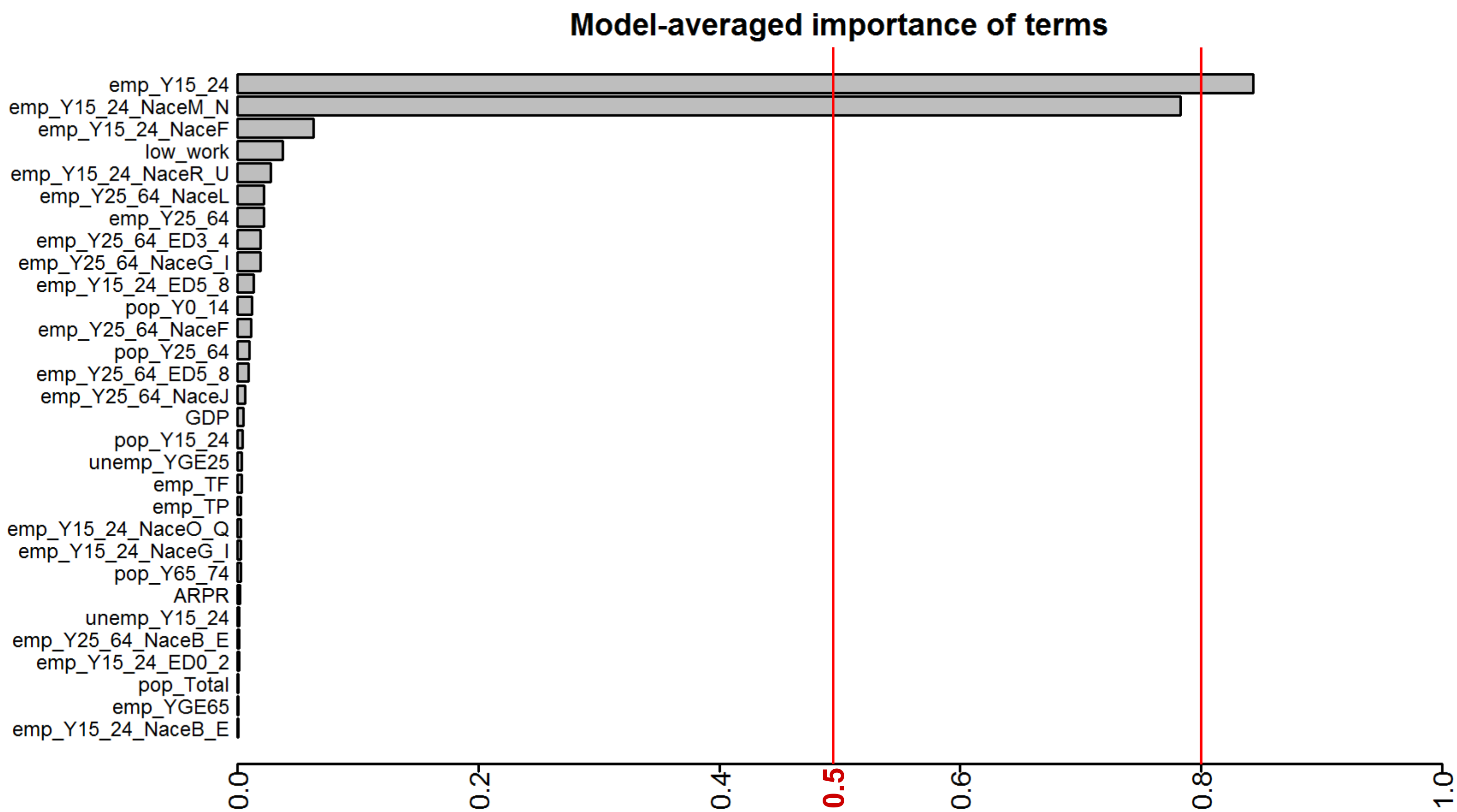
"ES-IT"
emp_Y15-24
pop_Y15-24

"IE" *
GVAgr
pop_Y0-14
pop_Y[65-74,GE75]

"LU" *
emp_Y25.64_ED5.8
GDP



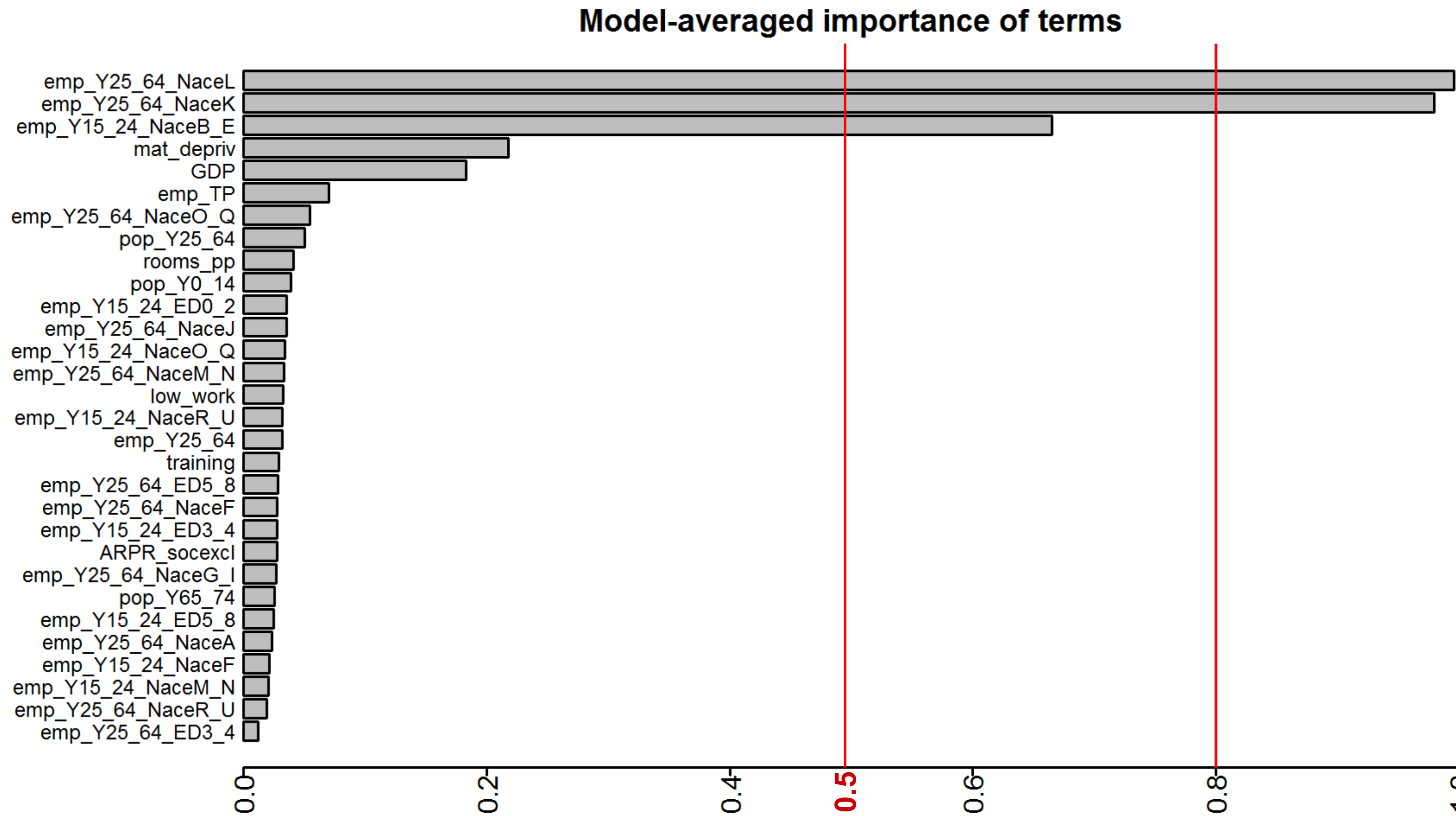
Results: Skills mismatch



mismatch = 100.6 – 3.2*emp_Y15-24 + 6.6*emp_Y15-24_NaceM-N

R^2 = 0.95

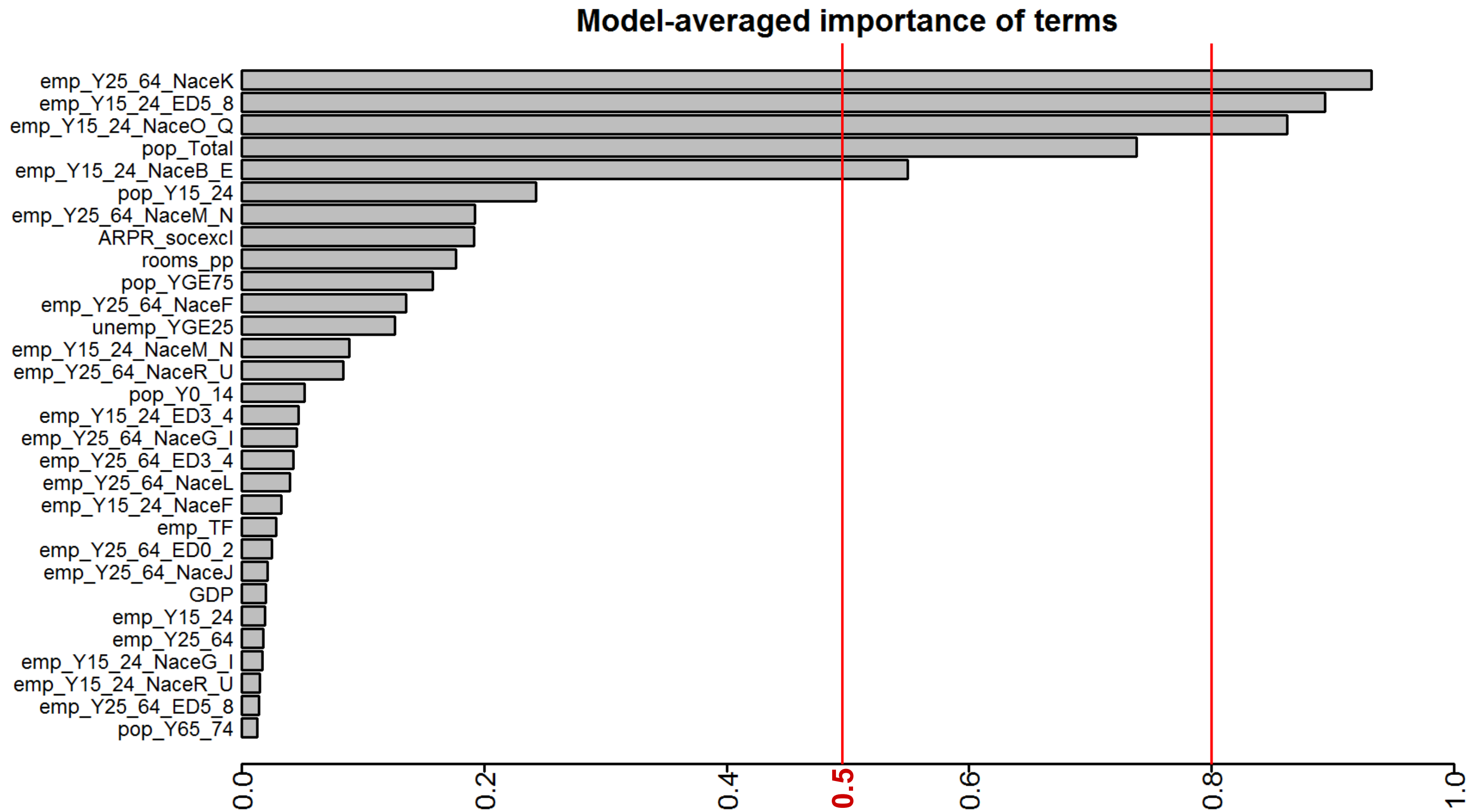
Results: LMkt Mobility



lmktm = 1.8 + 4.6*emp_Y25-64_NaceK + 10.0*emp_Y25-64_NaceL – 0.4*emp_Y15-24_NaceB-E

R^2 = 0.79

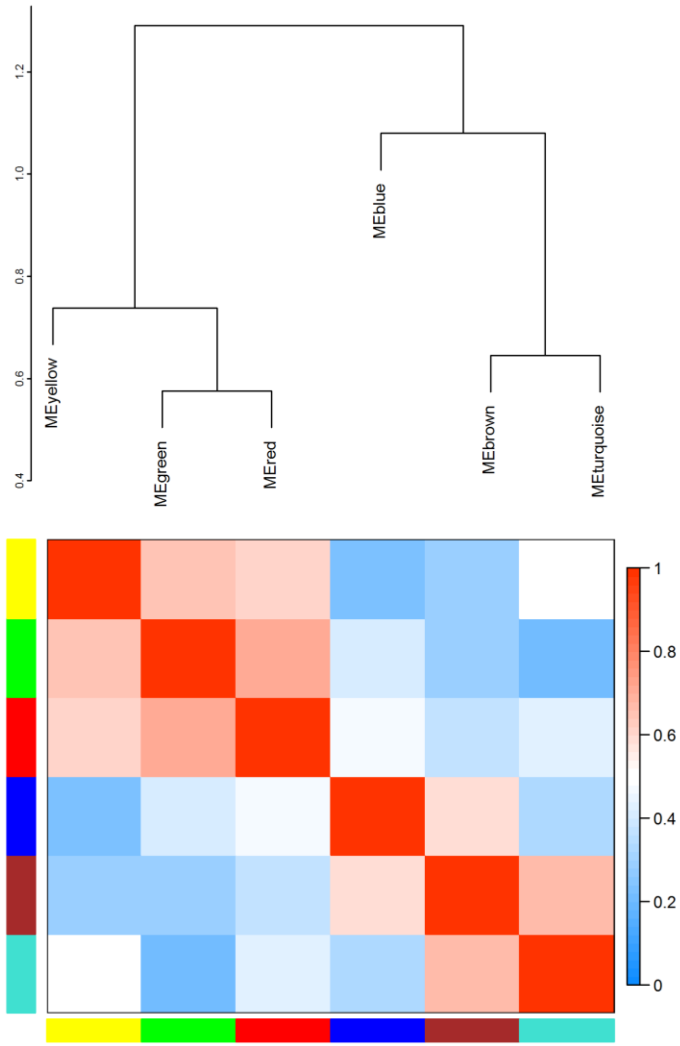
Results: Emigration



migrt = 0.52 + 0.18*emp_Y25-64_NaceK + 0.03*emp_Y15-24_ED5-8 – 0.03*emp_Y15-24_NaceO-Q – 5.80e^{-9} *pop_Total – 0.03*emp_Y15-24_NaceB-E

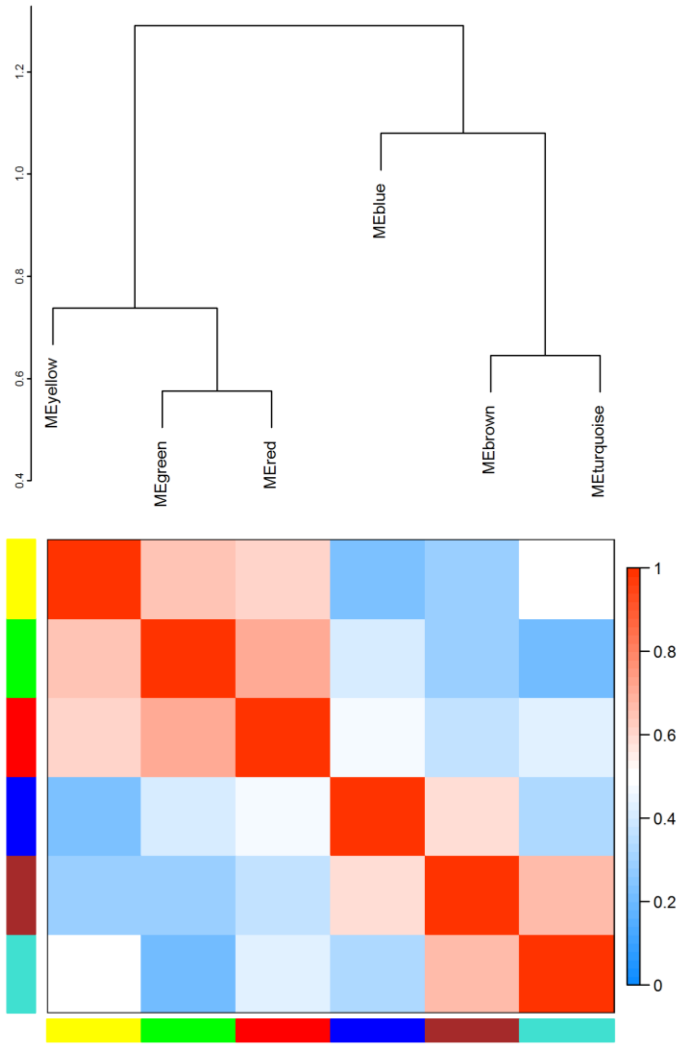
R^2 = 0.83

Results: WCNA



MEyellow	+ emp_Y[15-24]_Nace[G-I]
	+ unemp_Y[15-24,25-64]
MEgreen	**+ ilc_ARPR**
	+ ilc_ARPR_socexcl
MEred	**+ pop_Y[GE75]**
	– pop_Y[15-24]
MEblue	**+ emp_Y[25-64]_ED[3-4]**
	– emp_Y[25-64]_ED[0-2]
MEbrown	**+ emp_Y[15-24,25-64]**

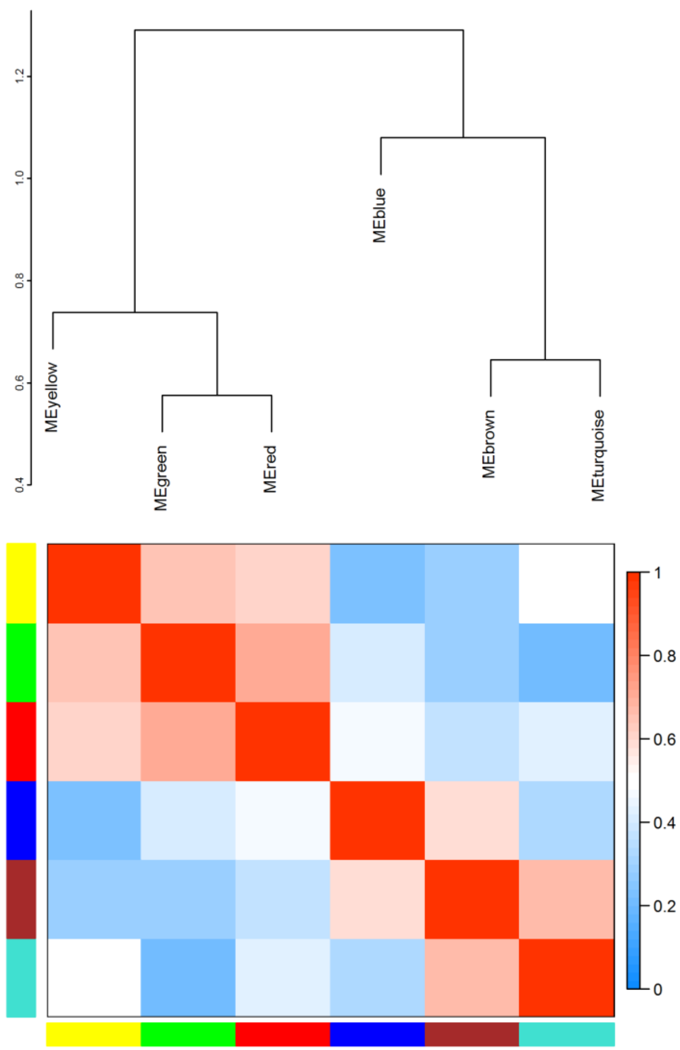
Results: WCNA



MEturquoise

- **+ ilc_rooms_pp**
- **+ earn_OC[1-5,7-9]_Nace[B-F,G-N]**
- **+ earn_OC[1-5,9]_Nace[P-S]**
- + emp_T[P]
- + emp_Y[25-64]_Nace[M-N,O-Q]
- **+ na_GDP**
- + training
- **– ilc_mat_depriv**

Results: WCNA



labels	description	mismatch	lmktm	migrt
MEyellow	Unemployment	0.36	–	–
MEgreen	Poverty	0.38	–	–
MEred	Ageing Population	–	–	-0.36
MEblue	Secondary Education (Employed Adults)	-0.38	–	-0.50
MEbrown	Employment	-0.69	0.35	–
MEturquoise	Earnings	–	0.59	–

Conclusions I

- **LMkt Attract** is able to form consistent clusters at NUTS0

- **LMkt Attract** can be reduced to 6 Eigenvariables: **Unemployment**, **Poverty**, **Ageing Population**, **Secondary Education (Employed Adults)**, **Employment** and **Earnings**

- **Skills Mismatch** decreases with **emp_Y15-24**, and increases with **emp_Y15-24_NaceM-N** (R^2=0.95,n=8)

- **Skills Mismatch** is negatively associated to **Employment** and **Secondary Education (Employed Adults)** and is positively associated to **Unemployment** and **Poverty**

Conclusions II

- **LMkt Mobility** increases with **emp_Y25-64_NaceL** and **emp_Y25-64_NaceK**, and decreases with **emp_Y15-24_NaceB-E** (R^2=0.79,n=25)
- **LMkt Mobility** is positively associated with **Employment** and **Earnings**
- **Emigration** increases with **emp_Y25-64_NaceK** and **emp_Y15-24_ED5-8**, and decreases with **pop_Total**, **emp_Y15-24_NaceO-Q** and **emp_Y15-24_NaceB-E** (R^2=0.83,n=28)
- **Emigration** is negatively associated with **Ageing Population** and **Secondary Education (Employed Adults)**

Acknowledge

Team:

Sónia Quaresma

João Lopes

Marco Moura

Institution:

Data:

https://github.com/jsollari/EUhackathon2017

Thank you!

Bibliography

1. https://ec.europa.eu/commission/publications/skills-education-and-lifelong-learning-european-pillar-social-rights_en

2. CEDEFOP (2015) "Skills, qualifications and jobs in the EU: the making of a perfect match? "

3. Council Decision (EU) 2015/1848 of 5 October 2015

4. Reynolds et al. (1992) "Clustering rules: A comparison of partitioning and hierarchical clustering algorithms" J Math. Model. Algorithms

5. Langfelder and Horvath (2008) "WGCNA: an R package for weighted correlation network analysis" BMC Bioinformatics

Tools

R libraries

car - Companion to Applied Regression

caret - Classification and Regression Training

cluster - Finding Groups in Data: Cluster Analysis Extended

glmulti - Model selection and multimodel inference made easy

MASS - Support Functions and Datasets for MASS

nnet - Feed-Forward Neural Networks and Multinomial Log-Linear Models

sna - Tools for Social Network Analysis

WGCNA - Weighted Correlation Network Analysis

Metadata: ISCED 11

label	description
ED0-2	Less than primary, primary and lower secondary education (levels 0-2)
ED3_4	Upper secondary and post-secondary non-tertiary education (levels 3 and 4)
ED5-8	Tertiary education (levels 5-8)

Metadata: ISCED-F 13

label	description
F00	Generic programmes and qualifications
F01	Education
F02	Arts and humanities
F03	Social sciences, journalism and information
F04	Business, administration and law
F05	Natural sciences, mathematics and statistics
F06	Information and Communication Technologies
F07	Engineering, manufacturing and construction
F08	Agriculture, forestry, fisheries and veterinary
F09	Health and welfare
F10	Services

Metadata: ISCO-08

label	description
OC1-5	Non manual workers
OC1	Managers
OC2	Professionals
OC3	Technicians and associate professionals
OC4	Clerical support workers
OC5	Service and sales workers
OC6-8	Skilled manual workers
OC6	Skilled agricultural, forestry and fishery workers
OC7	Craft and related trades workers
OC8	Plant and machine operators and assemblers
OC9	Elementary occupations
OC0	Armed forces occupations

Metadata: NACE Rev. 2

label	description
A	Agriculture, forestry and fishing
B-E	Industry (except construction)
B-F	Industry and construction
F	Construction
G-I	Wholesale and retail trade, transport, accommodation and food service activities
G-N	Services of the business economy
J	Information and communication
K	Financial and insurance activities
L	Real estate activities
M-N	Professional, scientific and technical activities; administrative and support service activities
O-Q	Public administration, defence, education, human health and social work activities
P-S	Education; human health and social work activities; arts, entertainment and recreation; other service activities
R-U	Arts, entertainment and recreation; other service activities; activities of household and extra-territorial organizations and bodies

Data: LMkt attractiveness

dataset	description	year	NUTS	units
demo_r_d2jan	Population	2014	NUTS 2	NR
earn_ses_hourly	Structure of earnings: hourly earnings	2014	NUTS 0	MN_PPS
educ_uoe_fine06	Total public expenditure on education	2013	NUTS 0	PC_GDP
ilc_li41	At-risk-of-poverty rate	2014	NUTS 2	PC_POP
ilc_lvhl21	People living in households with very low work intensity	2014	NUTS 2	PC_YLE60
ilc_lvho04n	Average number of rooms	2014	NUTS 2	AVG
ilc_mddd21	Severe material deprivation rate	2014	NUTS 2	PC_POP
ilc_peps11	People at risk of poverty or social exclusion	2014	NUTS 2	PC_POP
lfst_r_lfe2eedu	Employment by educational attainment level (ISCED 11)	2014	NUTS 2	THS
lfst_r_lfe2eftpt	Employment by full-time/part-time	2014	NUTS 2	THS
lfst_r_lfe2emp	Employment	2014	NUTS 2	THS
lfst_r_lfe2en2	Employment by economic activity (NACE Rev. 2)	2014	NUTS 2	THS
lfst_r_lfu3pers	Unemployment	2014	NUTS 2	THS
nama_10r_2gdp	Gross domestic product	2014	NUTS 2	PPS_HAB
nama_10r_2gvagr	Real growth rate of regional gross value added	2014	NUTS 2	PCH_PRE
nama_10r_2hhinc	Income of households	2013	NUTS 2	PPCS_HAB
trng_lfse_04	Participation rate in education and training (last 4 weeks)	2014	NUTS 2	PC_Y25-64

Data: LMkt attractiveness

variable	description	NUTS	units
ilc_ARPR	At-risk-of-poverty	NUTS 2	PC_POP
ilc_ARPR_socexcl	At-risk-of-poverty or social exclusion	NUTS 2	PC_POP
ilc_low_work	Very low work intensity	NUTS 2	PC_POP_YLE60
ilc_mat_depriv	Severe material deprivation	NUTS 2	PC_POP
ilc_rooms_pp	Number of rooms per person	NUTS 2	AVG
earn_OC[**titles**]_Nace[**sector**]	Earning by ISCO-08 title and NACE Ver. 2 sector	NUTS 0	MN_PPS
emp_[**contract**]	Employment by work contract	NUTS 2	PC_POP_YGE15
emp_Y[**age**]	Employment by age	NUTS 2	PC_POP_Y[**age**]
emp_Y[**age**]_ED[**level**]	Employment by age and ISCED 11 level	NUTS 2	PC_EMP_Y[**age**]
emp_Y[**age**]_Nace[**sector**]	Employment by age and NACE Ver. 2 sector	NUTS 2	PC_EMP_Y[**age**]
unemp_Y[**age**]	Unemployment by age	NUTS 2	PC_POP_Y[**age**]
expend_ED5-8	Public expenditure on education	NUTS 0	PC_GDP
na_disp_income	Disposable income	NUTS 2	PPCS_HAB
na_GDP	Gross Domestic Product	NUTS 2	PPS_HAB
na_GVAgr	Gross Value Added growth	NUTS 2	PCH_PRE
pop_Total	Population	NUTS 2	NR
pop_Y[**age**]	Population by age	NUTS 2	PC_POP
training	Participation in education and training	NUTS 2	PC_POP_Y25-64

Data: LMkt indicators

dataset	description	year	NUTS	units
educ_uoe_grad02	Graduates	2014	NUTS 0	NR
jvs_a_nace2	Job vacancies: annual data	2014	NUTS 0	NR
lfso_14leeow	Labour force	2014	NUTS 0	THS
migr_emi2	Emigration	2014	NUTS 0	NR

Data: LMkt indicators

variable	description	NUTS	units
mismatch	Skills mismatch	NUTS 2	DIST
lmktm_Total	Labour market mobility	NUTS 0	PC_EMP
lmktm_OC[**titles**]	Labour market mobility by ISCO-08 title	NUTS 0	PC_EMP_OC[**title**]
lmktm_ED[**level**]	Labour market mobility by ISCED 11 level	NUTS 0	PC_EMP_ED[**level**]
migrt_Total	Migration	NUTS 2	AVG
migrt_Y[**age**]	Migration by age	NUTS 0	PC_POP_Y[**age**]

Methods: details

- <u>Network Analysis</u>

  similarity: additive inverse of the weighted Euclidean distance

  transformation: min-max transformation

  edge threshold = {0.65, 0.8, 0.95}

  algorithm: "Fruchterman–Reingold" algorithm

- <u>Partition-around-medoids (PAM)</u>

  distance: weighted Euclidean distance

  k = {10, 21, 25}

- <u>Over-representation analysis (ORA)</u>

  num2cat: $x \in P_{0\text{-}10}$ and $x \in P_{90\text{-}100}$

  p-value correction: none

  p-value cut-off = 0.05

Methods: details

- <u>Multinomial regression</u>

remove NAs: column-wise (NA > 0.00) and row-wise (NA > 0.00)

remove predictors: i) r-between < 0.90; ii) r-within < [up-to 30 vars]

transformation: min-max transformation

max terms: hard threshold (i.e. nparams = nsubj – 1)

confidence set = 100

model level: only main effects

information criteria: AIC

models exploration: if nmods < 200000 exhaustive screening, else genetic algorithm

genetic algorithm: i) popsize = 100, mutrate = 10^{-3}, sexrate = 0.1, imm = 0.3, deltaM = 0.05, deltaB = 0.05, conseq = 5; ii) popsize = 200, mutrate = 10^{-2}, sexrate = 0.2, imm = 0.6, deltaM = 0.005, deltaB = 0.005, conseq = 10. Number replicates = 2.

model: multinomial logistic model via single-layer feed-forward neural networks

Methods: details

- <u>Multivariate regression</u>

remove NAs: column-wise (NA > 0.00) and row-wise (NA > 0.00)

remove predictors: i) r-between < 0.90; ii) r-within < [up-to 30 vars]

transformation: none

max terms: hard threshold (i.e. nparams = nsubj – 4)

confidence set = 100

model level: only main effects

information criteria: AICc

models exploration: if nmods < 200000 exhaustive screening, else genetic algorithm

genetic algorithm: i) popsize = 100, mutrate = 10^{-3}, sexrate = 0.1, imm = 0.3, deltaM = 0.05, deltaB = 0.05, conseq = 5; ii) popsize = 200, mutrate = 10^{-2}, sexrate = 0.2, imm = 0.6, deltaM = 0.005, deltaB = 0.005, conseq = 10. Number replicates = 2.

model: multivariate linear regression

Methods: details

- <u>Weighted correlation network analysis (WCNA)</u>

remove NAs: row-wise (NAs > 0.50) and col-wise (NAs > 0.15)

distance = Topological Overlap Matrix(|Spearman $r|^k$)

k-power transformation: k = min(k), when k > 0.9*k-best

minimum module size = 2

cluster splitting level = 4, where level $\in$ {1,2,3,4}

dynamic tree cut method = "hybrid"

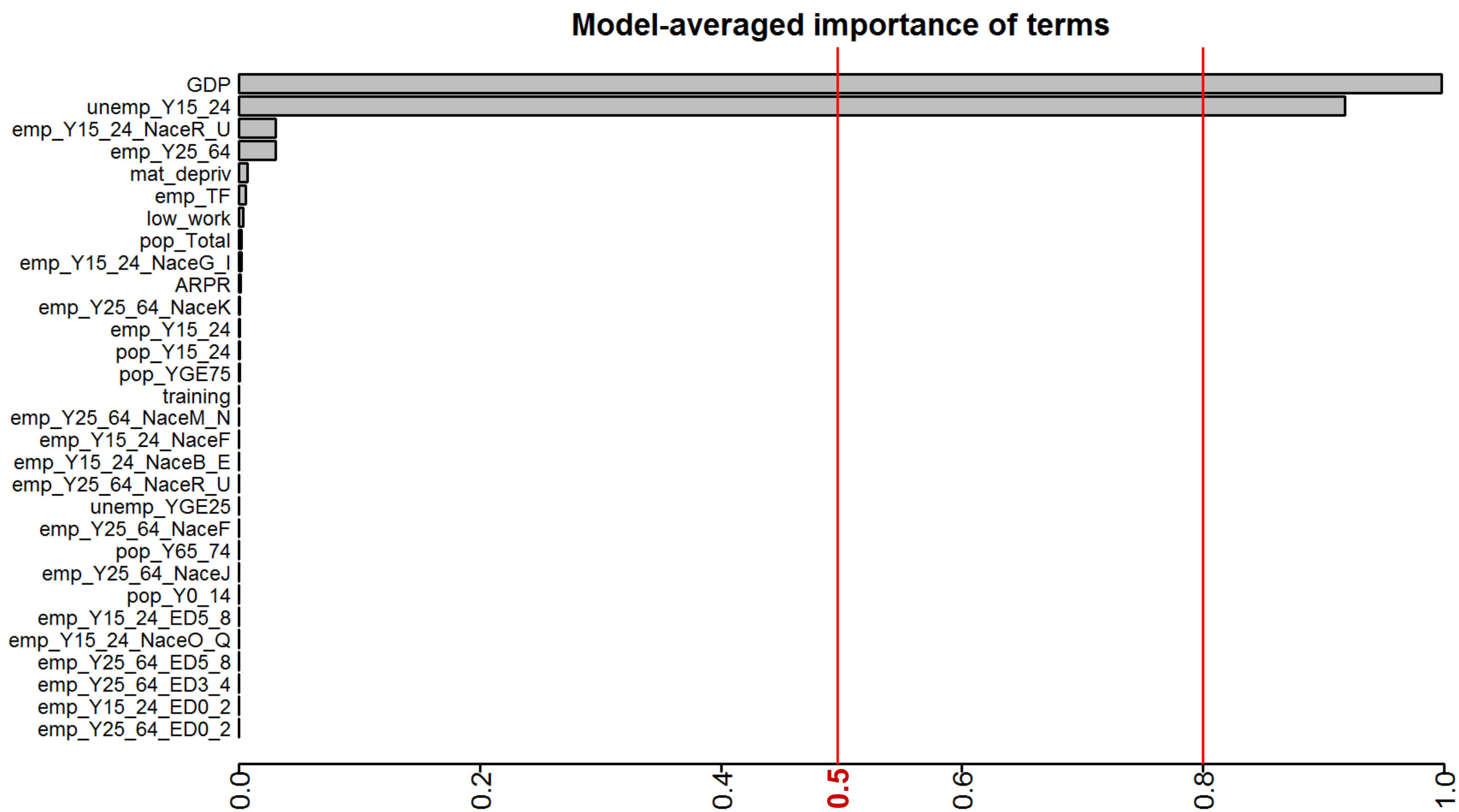
merge cut-off height = 0.2

distance between modules = 1 - |Spearman r (eigenvalues)|

association modules and variables: i) Spearman r (eigenvalues) for numeric variables; ii) $(R^2_{McFadden})^{½}$ for categorical variables

r cut-off = |0.30|

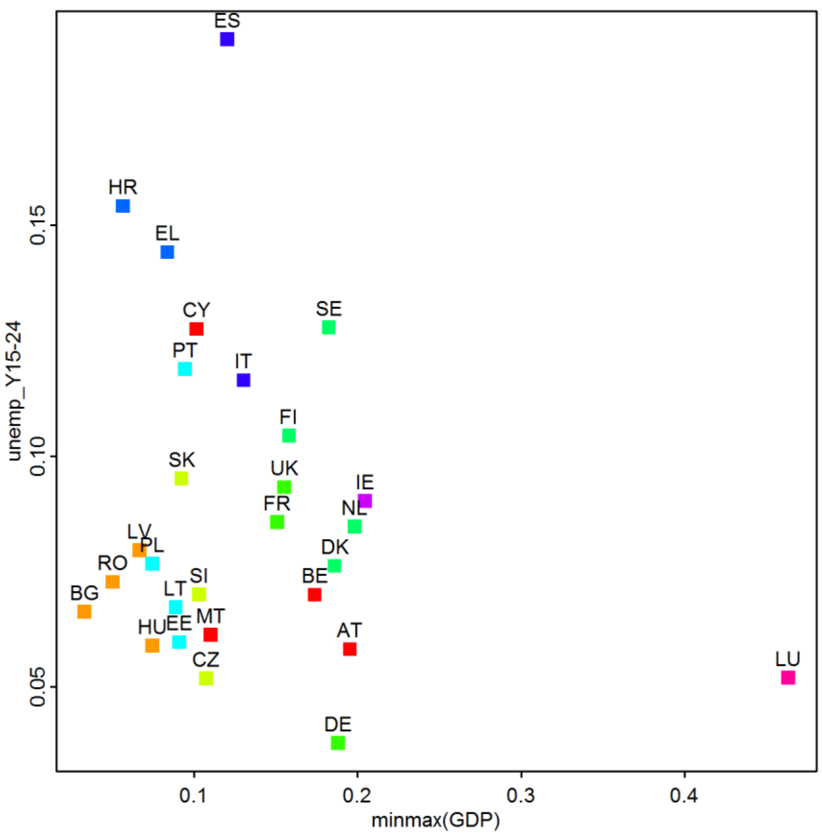
Results: LMkt Attractiveness

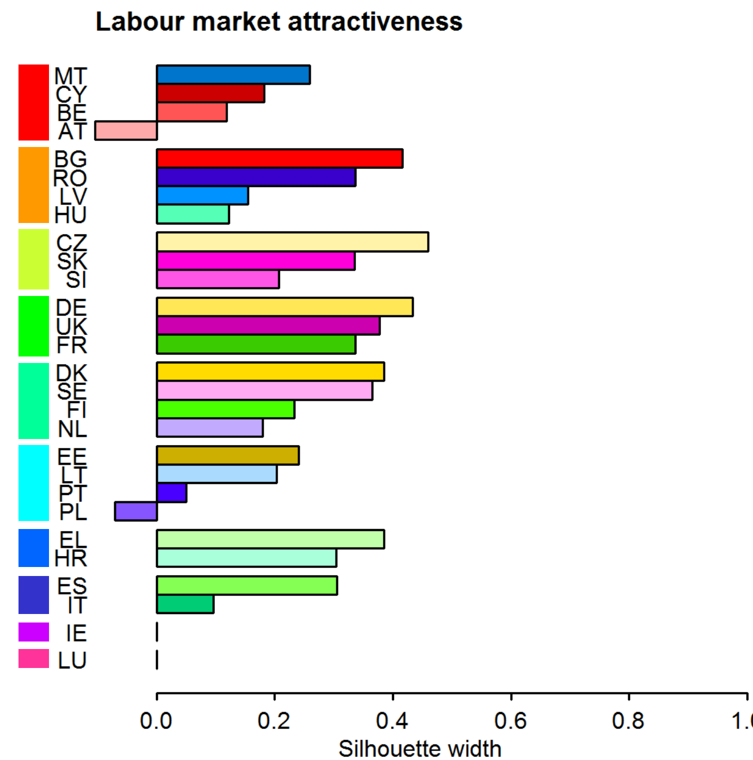


EU_groups ~ 1 + GDP + unemp_Y15-24

$R^2_{McFadden} = 0.84$, $R^2_{count} = 0.82$

Results: LMkt Attractiveness

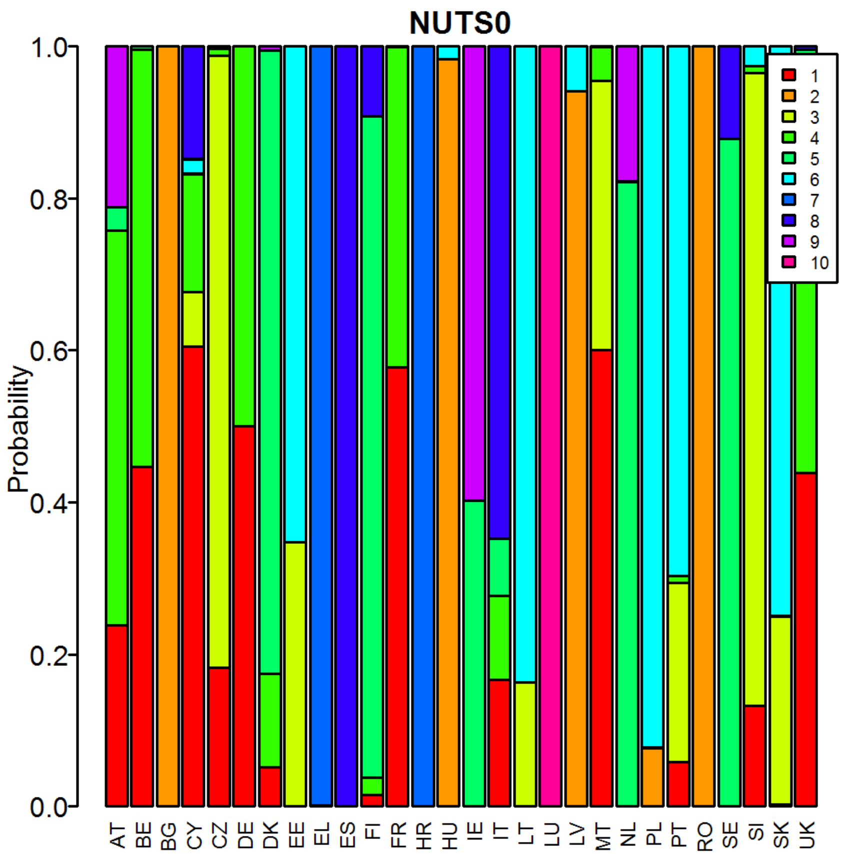


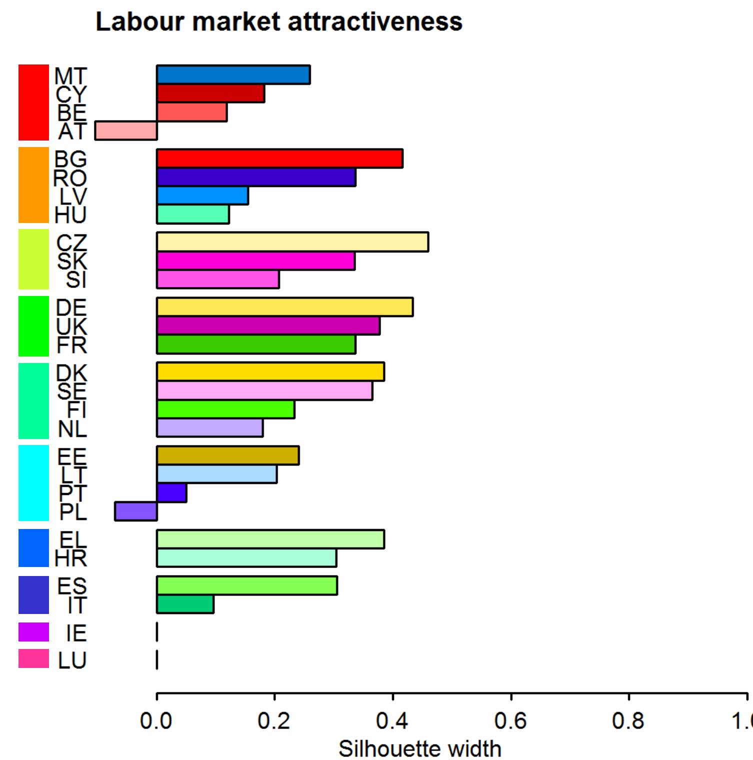


EU_groups ~ 1 + GDP + unemp_Y15-24

$R^2_{McFadden} = 0.84$, $R^2_{count} = 0.82$

Results: LMkt Attractiveness

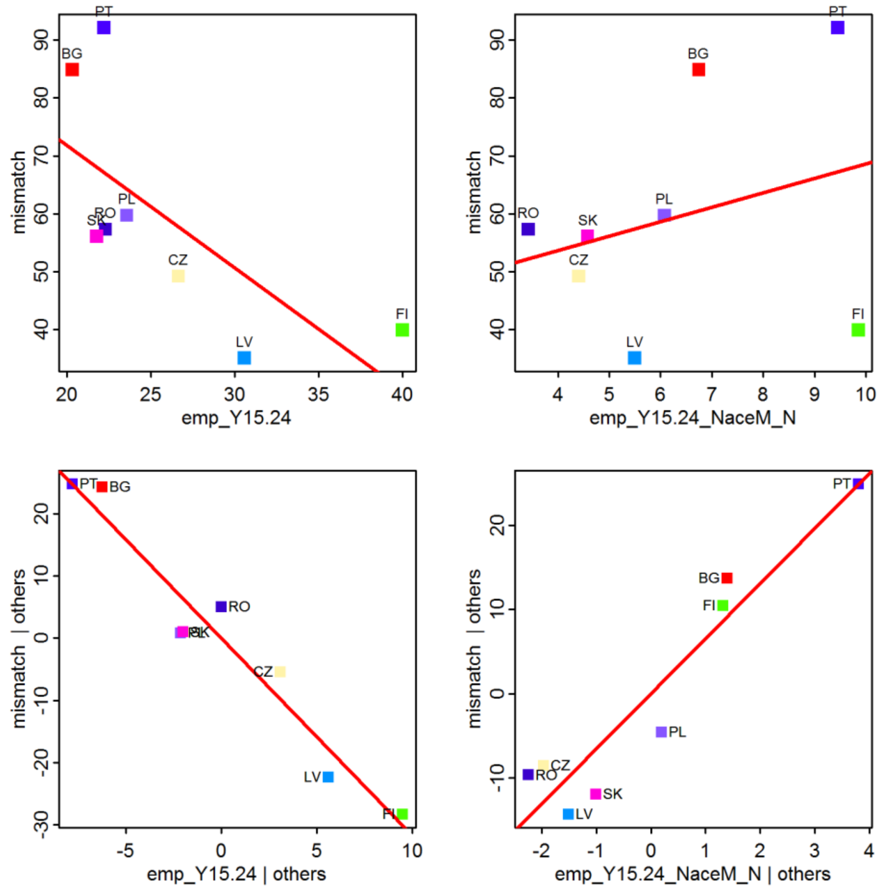




EU_groups ~ 1 + GDP + unemp_Y15-24

$\mathbf{R^2_{McFadden} = 0.84,\ R^2_{count} = 0.82}$

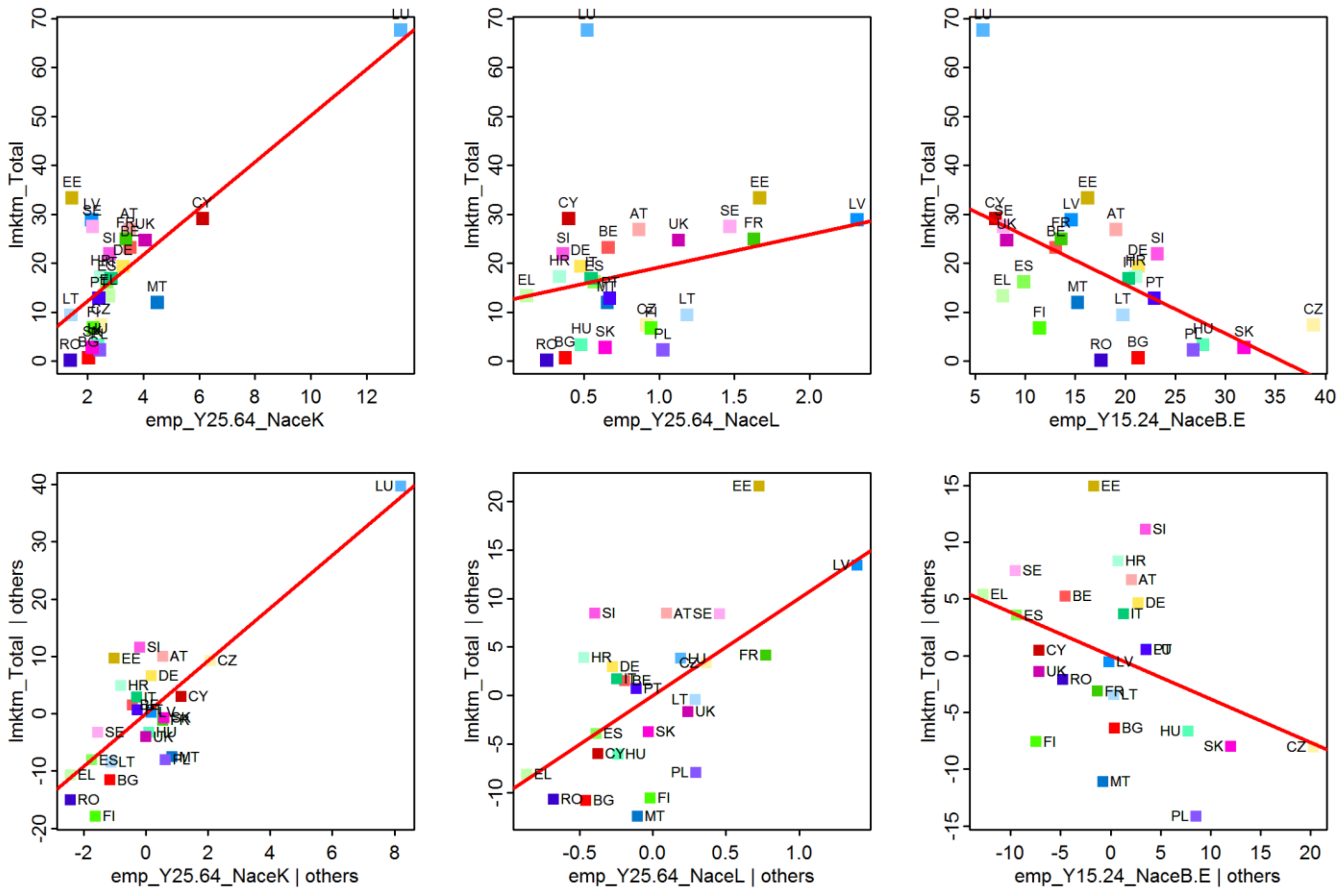
Results: Skills mismatch



mismatch = **100.6 – 3.2*emp_Y15-24 + 6.6*emp_Y15-24_NaceM-N**

R^2 = 0.95

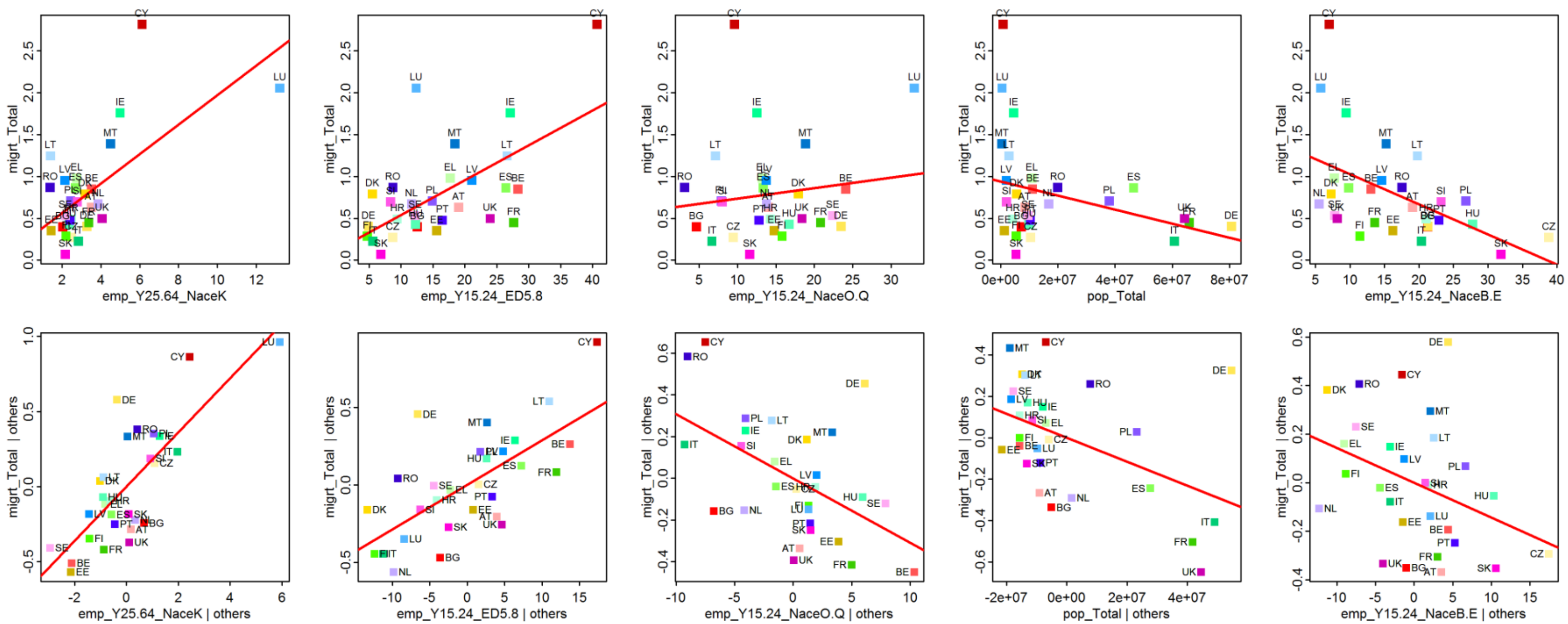
Results: LMkt Mobility



lmktm = 1.8 + 4.6*emp_Y25-64_NaceK + 10.0*emp_Y25-64_NaceL – 0.4*emp_Y15-24_NaceB-E

R^2 = 0.79

Results: Emigration



migrt = 0.52 + 0.18*emp_Y25-64_NaceK + 0.03*emp_Y15-24_ED5-8 – 0.03*emp_Y15-24_NaceO-Q – 5.80e^{-9} *pop_Total – 0.03*emp_Y15-24_NaceB-E

R^2 = 0.83