# Thesis Update

## The Determinants of Economic Growth in EU28 and Candidate Countries. A Regional Analysis.

Elia Di Gregorio

Vienna University of Economics and Business

September 2024

# Overview

# Introduction

**2004 "Big Bang" EU Enlargement**

- 10 Central and Eastern European (CEE) countries joined the EU.
- Marked a significant step in European integration.

**Current Challenges for the EU** e.g. Brexit, COVID-19 pandemic, Russian invasion of Ukraine, and inflation reinforced the importance of EU unity and accelerated accession efforts.

**Accession Efforts for Western Balkan Countries, Moldova, and Ukraine**

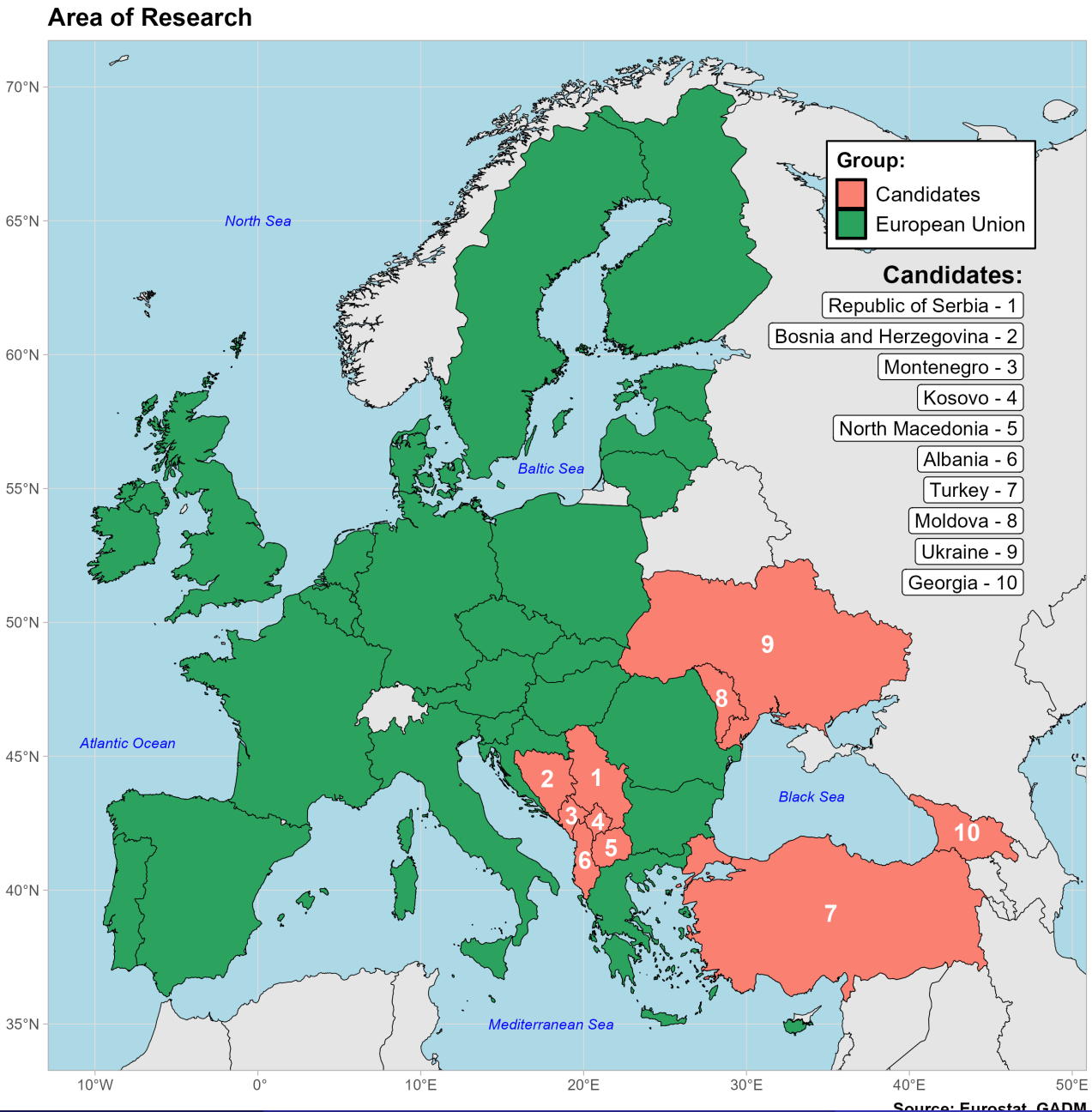
- But challenges include compliance with the *acquis communautaire*, military/ethnic disputes, and corruption.

# Paper Inspiration

Cuaresma et al. [2014]: *The Determinants of Economic Growth in European Regions.*

- Analyze determinants of regional economic growth in 255 European regions (1995-2005) using Bayesian Model Averaging (BMA).
  - **Income Convergence:** Strong convergence observed between regions of new EU member states in Central and Eastern Europe (CEE). Especially in regions with capital cities grow significantly faster.
  - **Human Capital:** Share of workers with higher education drives positive GDP growth.
  - **Spatial Spillovers:** Positive spatial spillovers observed, but spillovers through explanatory variables are not substantial.

# EU Enlargement



# Literature Review

**Regional Growth Determinants and BMA**

- The use of Bayesian Model Averaging (BMA) in growth models allows for addressing uncertainty by averaging across multiple models, ensuring robust identification of growth determinants.

**Convergence and Regional Disparities**

- The thesis explores the theory of convergence, where poorer regions or countries grow faster than wealthier ones, and how this process has been uneven across Europe, particularly in Central, Eastern, and Southeastern European (CESEE) regions.

**Policy Implications and Cohesion**

- Literature focusses on the impact of EU cohesion policy, particularly how it has influenced regional growth strategies in both Member States and Candidate countries, and its varying effectiveness in reducing regional disparities.

# Paper Contribution

**Replication and Expansion of Cuaresma et al. [2014]**

- Applies the same methodological framework to 2009-2019 data.

**Comprehensive Analysis**

- Expands analysis to include following EU Candidate countries:

1. Republic of Serbia
2. Bosnia and Herzegovina
3. Montenegro
4. Kosovo
5. Albania
6. North Macedonia
7. Turkey
8. Moldova

**Policy Relevance**

- Inform policies targeting disparities between Member States and Candidate countries.

# Regional Coverage

This study covers 265 EU and 36 candidate NUTS-2 [1]. For candidate countries not following this system I recurred to the population thresholds of the NUTS classification:

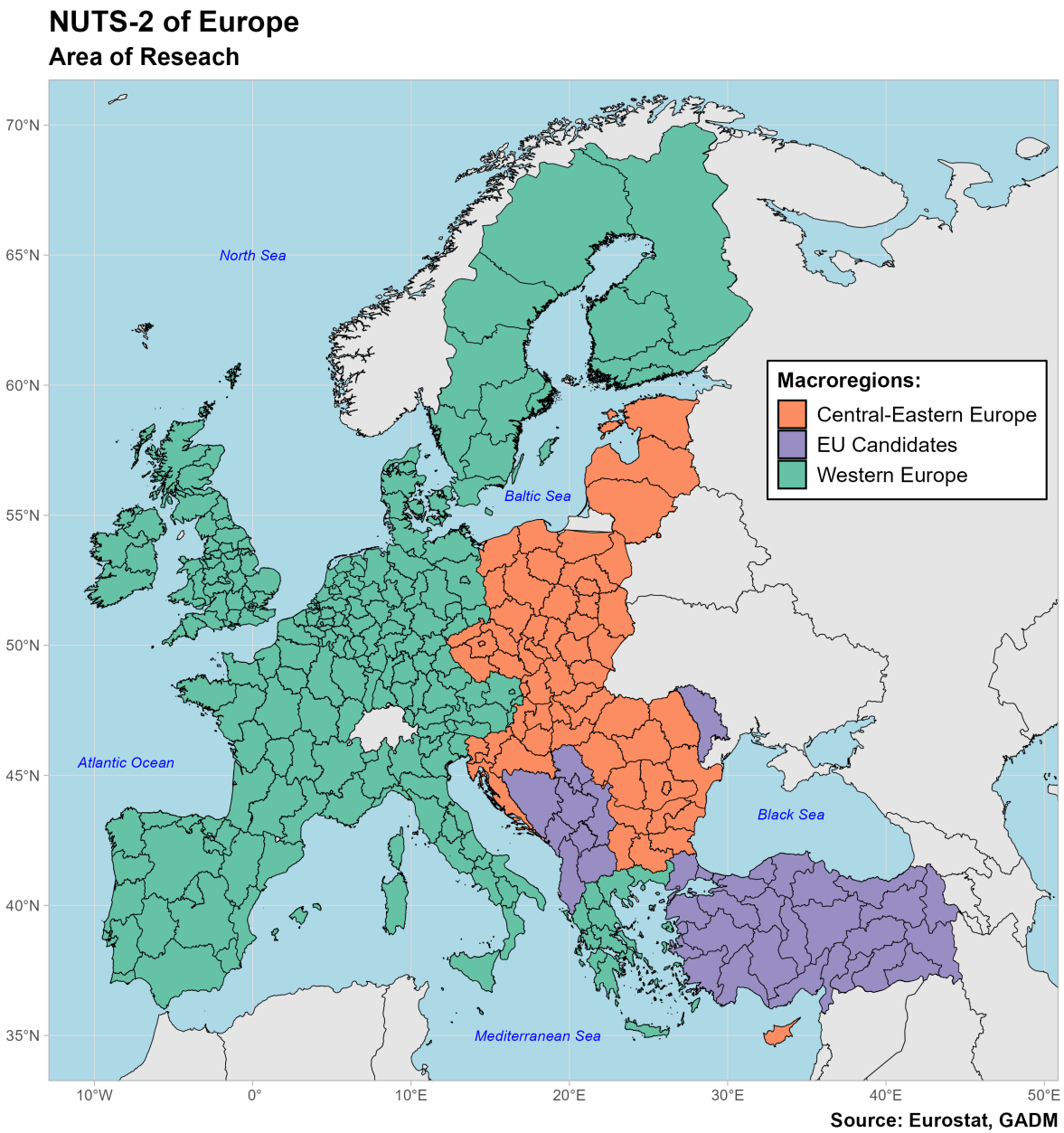
| NUTS Level | Min. Pop | Max. Pop |
|---|---|---|
| NUTS-1 | 3,000,000 | 7,000,000 |
| NUTS-2 | 800,000 | 3,000,000 |
| NUTS-3 | 150,000 | 800,000 |

Based on these thresholds and regional data availability, the candidate countries were divided into the following systems:

- **Single NUTS-2 regions** → Kosovo, North Macedonia, Montenegro, Albania, Bosnia and Herzegovina, Moldova.
- **Multiple NUTS-2 regions** → Serbia, Turkey.

[1] The 2021 NUTS Classification System was used with some exceptions

# NUTS-2 Regions



# Sources

**Sources:**

- EU Databases: ARDECO, Eurostat, ESPONS;
- World Bank;
- National Statistical Offices;
- WiiW;
- ILOSTAT.

# Variables

**Dependent variable**: GDP Growth (Base price EUR09)

**Explanatory Variables**:

1. GDP per capita (log),
2. Net Migration rate,
3. Labor Productivity (log),
4. Gross Wage in EUR,
5. Activity rate,
6. NEET rate,
7. Life expectancy,
8. Fertility rate,
9. Share of Population by Educational Attainment,
10. Employment rate,
11. Unemployment rate,
12. Investment rate,
13. Gross Fixed Capital Formation share (of GDP),
14. Sectoral components of GVA,
15. Output density,
16. Employment density ($km^2$),
17. Population density ($km^2$),
18. Distance from Bruxelles (km),
19. *Dummy variables and interactions.*

# Variables Breakdown

**Sectoral Components of GVA**[2]:

- Agriculture (NACE A)
- Industry (NACE B-E)
- Construction (NACE F)
- Services (NACE G-N)
- Public Sector (NACE O-U)

**Educational Attainment**[3]**:**

- Basic Education (ISCED 0-2)
- Secondary Education (ISCED 3-4)
- Tertiary Education (ISCED 5-8)

**Dummy Variables:**

- CEE/Candidate Flag
- Capital Region
- Island Region
- Objective 1 Region
- Border Region

**Interaction Terms:** CEE/Candidate $\times$

- Capital region
- Population with tertiary education
- Sectoral components of GVA

[2]Based on NACE Rev. 2.
[3]Based on ISCED 2011.

# Basian Model Averaging

- BMA addresses model uncertainty by averaging across models, rather than selecting a single "best" model,
- it prevents issues like overfitting and multicollinearity,
- BMA accounts for the uncertainty in model selection and estimates the posterior distribution as a weighted average across models → models are weighted by their ability to explain the data.

- Key model specifications:
  1. Cross-sectional baseline model for regions.
  2. Baseline model expanded with fixed effects model to control for unobserved heterogeneity.
  3. Baseline adjusted by Spatial Autoregressive (SAR) model to capture spatial spillover effects.

# Econometric Model

The 3 BMA Model specification as in Cuaresma et al. [2014] which can all be nested within a general SAR model of the form:

$$\gamma = \alpha\iota_N + \rho W\gamma + X_K\vec{\beta}_k + \epsilon \tag{1}$$

- $\gamma$ is an N-dimensional column vector of stacked growth rates of income per capita for N regions;
- $\iota_N$ is an N-dimensional column vector of ones;
- $X_K = (x_1, \ldots, x_k)'$ is a matrix whose columns are stacked data for $K$ explanatory variables;
- $\vec{\beta}_k = (b_1, \ldots, b_k)'$ is the k-dimensional parameter vector corresponding to the variables in $X_K$;
- $W$ first-order queen contiguity matrix with inverse distance weights;
- $\rho$ is a scalar indicating the degree of spatial autocorrelation;
- $\epsilon$ is an error term which may contain country-specific fixed effects.

# Bayes' Theorem

- BMA averages over $2^K$ models which can be denoted as $M_j$ for $j = 1, ..., K$ with weights based on posterior model probabilities (PMPs).
- PMP are derived using Bayes' Theorem:

$$p(M_j|\mathcal{D}) = \frac{p(\mathcal{D}|M_j)p(M_j)}{p(\mathcal{D})} = \frac{p(\mathcal{D}|M_j)p(M_j)}{\sum_{j=1}^{K} p(\mathcal{D}|M_j)p(M_j)}$$

where

  - $p(M_j)$ is the prior on model $M_j$
  - $p(\mathcal{D}|M_j)$ is the likelihood of model $M_j$
  - $p(\mathcal{D})$ is the probability of $\mathcal{D}$ integrated over the model space

# Posterior Distribution

- Given the posterior model probabilities, we can write the posterior distribution of coefficient $\beta_k$ as

$$p(\beta_k|\mathcal{D}) = \sum_{j=1}^{K} p(\beta_k|M_i, \mathcal{D})p(M_i|\mathcal{D})$$

- The posterior distribution of the estimator of $\beta_k$ (given $\mathcal{D}$) is the average of its posterior distribution under each model, weighted by its posterior model probability.

# Prior Structure

For **Parameter Prior**: Zellner's g-prior, which can be written as:

$$p(\beta_k | \underbrace{(\alpha, \rho, \sigma, M_j)}_{g}) \sim \mathcal{N}(0, \sigma^2 g(X_k' X_k)^{-1}),$$

- I used the benchmark prior suggested by Fernandez et al. [2001], i.e $g = max(N, K^2)$, where $N$ is the number of observations and $K$ is the number of covariates.

For **Model Prior**: binomial-beta hyperprior on the a priori inclusion probability with prior expected value of the model size prior set to $K/2$.

# Markov Chain Monte Carlo

The $MC^3$ allows to identify the most important models and make reasonable predictions without the need to explore all $2^K$ models. Birth-death sampler used to add or drop covariates systematically:

$$p_{i,j} = \min\left(1, \frac{p(M_j|\mathcal{D})}{p(M_i|\mathcal{D})}\right)$$

- This sampler starts at model $M_i$ with a posterior model probability of $p(M_i|\mathcal{D})$.
- One of the $k$ potential covariates in $X_k$ is randomly chosen.
- If the covariate is part of $M_i$, it is dropped ("dies"), otherwise it is added ("born") to create another model $M_j$.
- The sampler now switches from $M_i$ to $M_j$ with the probability $p_{i,j}$.

# Spatial Weight Matrices

- Multiple spatial weight matrices are used to capture different spatial dependencies in the SAR model:
  1. Inverse distance matrix: $W_{ij} = d_{ij}^{-\phi}$ with $\phi = \{1, 2\}$
  2. First-order and second-order queen-contiguity matrices: Neighbors and neighbors-of-neighbors.
  3. k-nearest neighbors (k-NN): Defines spatial relationships by closest $k = 5$ neighbors.
- This approach ensures robustness by integrating over multiple spatial structures.

# Additional

The whole research was carried out on R and relied on the following packages for the BMA: `bms` and `spatBMS` developed by Feldkircher and Zeugner [2015].

Variables with $PIP > 0.5$ are considered robust following Kass and Raftery [1995]:

- Weak: 50–75%
- Substantial: 75–95%
- Strong: 95–99%
- Decisive: 99% or higher

## Strong Heredity Principle

The $MC^3$ runs with 10 million draws and a 3 million iteration burn-in period, and is adjusted to satisfy the *strong heredity principle*: interaction terms are only included in models when their corresponding main effects (i.e., the individual interacted variables) are also present.

# Datasets

I conducted the analysis on five distinct datasets to evaluate the consistency of results across different temporal and regional perspectives, and to control for potential variations in economic dynamics and structural differences among regions.

- Dataset with pool regions across 11 years,
- Dataset with cross regions: 2009-2019,
- Dataset with 2 panel periods: 2009-2014 & 2015-2019,
- Dataset with only CESEE cross regions: 2009-2019,
- Dataset with only EU cross regions: 2009-2019.

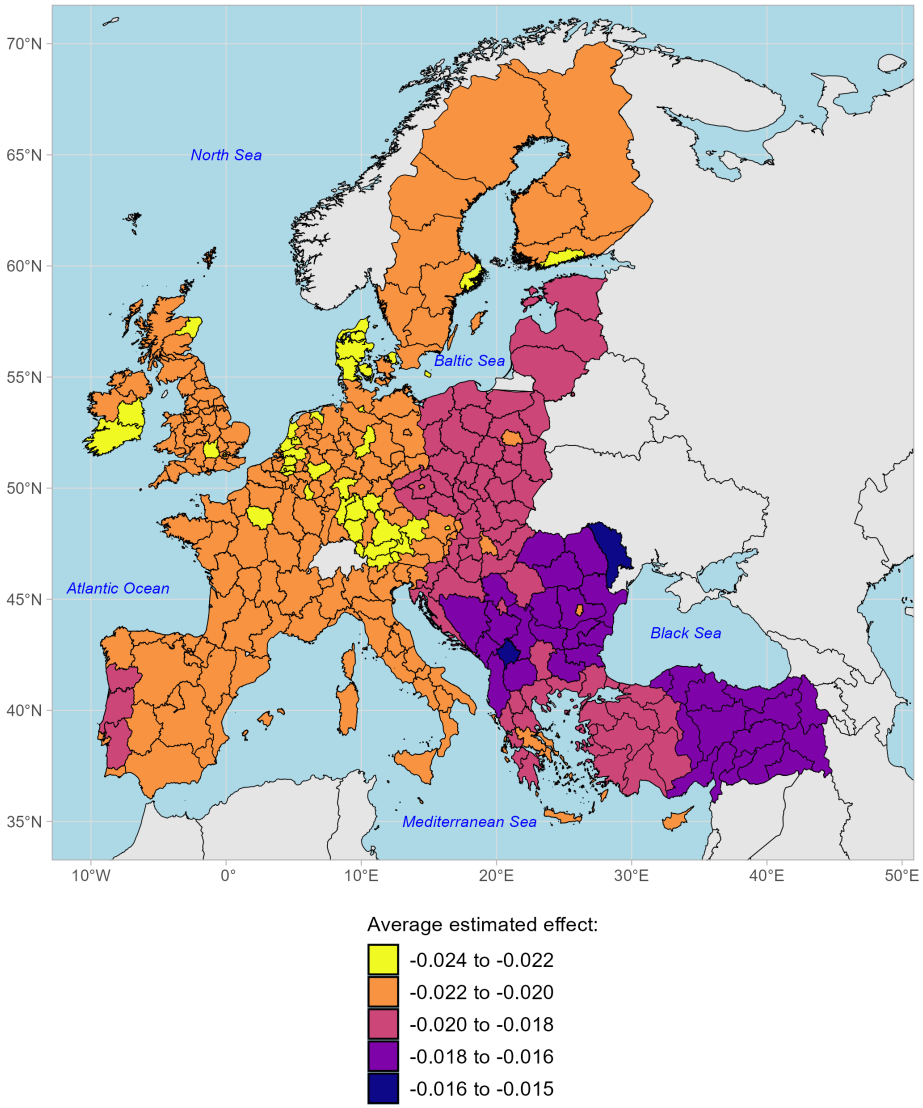
# Between Countries Baseline Model

| | *Model 1* | | | *Model 2* | | | *Model 3* | | |
|---|---|---|---|---|---|---|---|---|---|
| **Variable** | **PIP** | **PM** | **PSD** | **PIP** | **PM** | **PSD** | **PIP** | **PM** | **PSD** |
| Unemployment Rate | 1.000 | -0.118 | 0.013 | 0.999 | -0.073 | 0.017 | - | - | - |
| NEET Share | 0.999 | 0.073 | 0.013 | 0.812 | 0.042 | 0.025 | - | - | - |
| GVA Construction | 0.999 | -0.153 | 0.031 | 0.999 | -0.147 | 0.034 | 0.998 | -0.149 | 0.033 |
| GDP per Capita | 0.998 | -0.017 | 0.005 | - | - | - | - | - | - |
| Population with Tertiary Education | 0.985 | 0.055 | 0.010 | 0.989 | 0.050 | 0.010 | 1.000 | 0.046 | 0.009 |
| Migration Rate | 0.834 | -0.268 | 0.145 | 0.990 | -0.386 | 0.096 | - | - | - |
| GFCF Share | 0.580 | 0.035 | 0.033 | - | - | - | - | - | - |
| Labor Productivity | 0.569 | 0.005 | 0.005 | 0.504 | 0.006 | 0.007 | 1.000 | 0.018 | 0.004 |
| Capital | - | - | - | 0.792 | -0.006 | 0.004 | 0.999 | -0.002 | 0.003 |
| Distance from Bruxelles | - | - | - | - | - | - | 0.982 | 0.000 | 0.000 |
| *CEE/Candidates - Dummy interactions* | | | | | | | | | |
| Candidates | | | | 1.000 | 0.039 | 0.004 | 1.000 | 0.107 | 0.011 |
| CEE | | | | 1.000 | 0.023 | 0.003 | 1.000 | -0.010 | 0.010 |
| Candidates× GVA Industry | | | | | | | 1.000 | -0.154 | 0.020 |
| Candidates× Capital | | | | | | | 0.985 | -0.025 | 0.007 |
| CEE× Population with Tertiary Education | | | | | | | 0.999 | 0.142 | 0.023 |
| CEE× Capital | | | | | | | 0.999 | -0.026 | 0.005 |
| Share of posterior probabilities - Best model | | 0.16 | | | 0.14 | | | 0.10 | |
| Share of posterior probabilities - Top 25 models | | 0.73 | | | 0.69 | | | 0.41 | |
| Share of posterior probabilities - Top 50 models | | 0.80 | | | 0.77 | | | 0.49 | |
| **Corr PMP** | | 0.9996 | | | 1.0000 | | | 0.9999 | |

# Distribution Coefficients

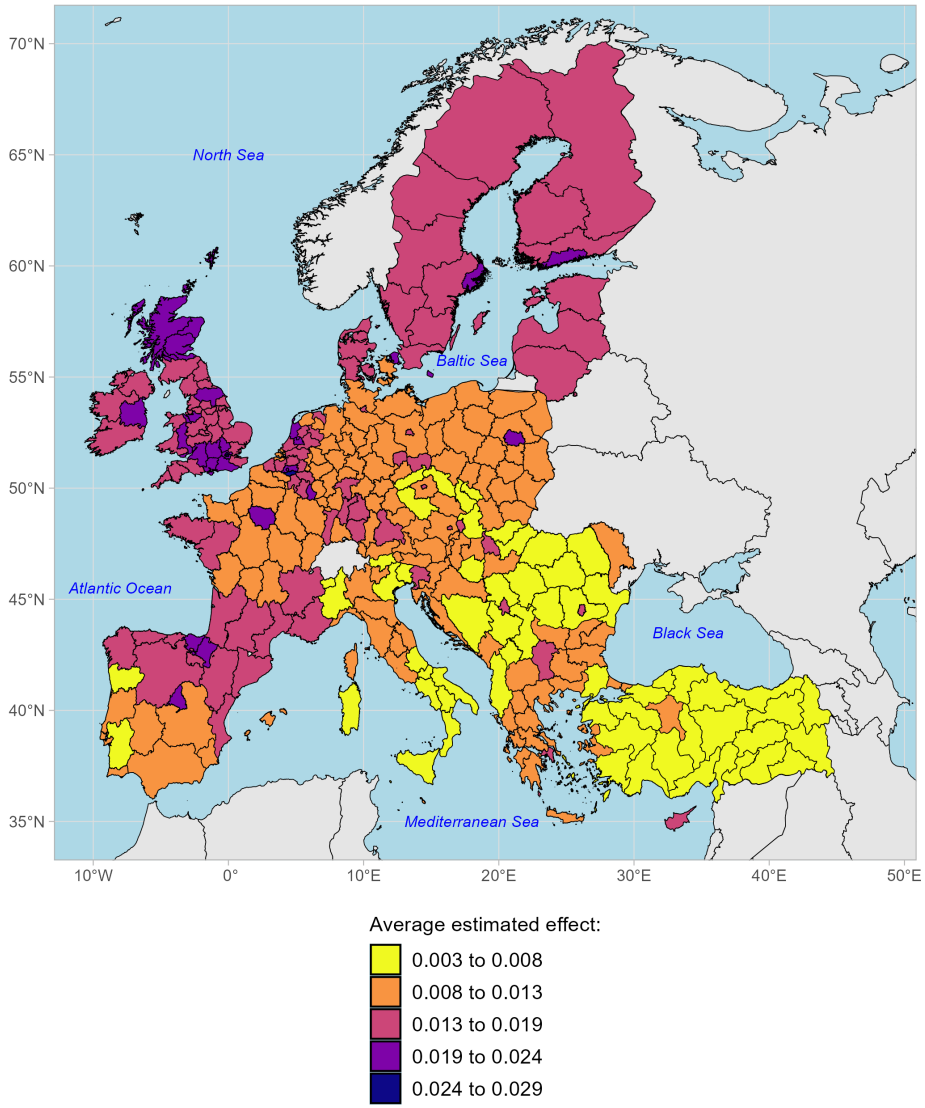
**Spatial Distribution PMs**

**Initial Income**




Average estimated effect:
- -0.024 to -0.022
- -0.022 to -0.020
- -0.020 to -0.018
- -0.018 to -0.016
- -0.016 to -0.015

**Human Capital**




Average estimated effect:
- 0.003 to 0.008
- 0.008 to 0.013
- 0.013 to 0.019
- 0.019 to 0.024
- 0.024 to 0.029

# Within Countries Baseline Model

| | *Model 1* | | | *Model 2* | | |
|---|---|---|---|---|---|---|
| **Variable** | **PIP** | **PM** | **PSD** | **PIP** | **PM** | **PSD** |
| Investment Rate | 1.000 | 1.355 | 0.203 | 1.000 | 1.419 | 0.240 |
| Gross Fixed Capital Formation share | 1.000 | -1.517 | 0.228 | 1.000 | -1.592 | 0.268 |
| NEET Share | 0.999 | 0.111 | 0.023 | 0.771 | 0.050 | 0.031 |
| GVA Construction | 0.999 | -0.259 | 0.038 | 0.604 | -0.138 | 0.123 |
| Population with Basic Education | 0.938 | -0.073 | 0.023 | 0.980 | -0.079 | 0.017 |
| GVA Agriculture | 0.814 | 0.064 | 0.036 | - | - | - |
| Unemployment Rate | 0.625 | -0.043 | 0.037 | - | - | - |
| *Dummy interactions* | | | | | | |
| Candidates×GDP per Capita | | | | 0.999 | -0.049 | 0.008 |
| CEE×GDP per Capita | | | | 0.541 | 0.019 | 0.022 |
| Candidates×GVA Services | | | | 0.998 | 0.128 | 0.091 |
| CEE×GVA Construction | | | | 0.706 | -0.273 | 0.212 |
| Share of posterior probabilities - Best model | | 0.22 | | | 0.11 | |
| Share of posterior probabilities - Top 25 models | | 0.74 | | | 0.45 | |
| Share of posterior probabilities - Top 50 models | | 0.81 | | | 0.54 | |
| **Corr PMP** | | 1.0000 | | | 0.9998 | |

# SAR Baseline Model

| | *Model 1* | | | *Model 2* | | | *Model 3* | | |
|---|---|---|---|---|---|---|---|---|---|
| **Variable** | **PIP** | **PM** | **PSD** | **PIP** | **PM** | **PSD** | **PIP** | **PM** | **PSD** |
| GVA Construction | 0.999 | -0.231 | 0.035 | 0.999 | -0.228 | 0.035 | 1.000 | -0.205 | 0.032 |
| NEET Share | 0.993 | 0.081 | 0.020 | 0.993 | 0.082 | 0.020 | 0.717 | 0.029 | 0.022 |
| Population with Tertiary Education | 0.969 | 0.044 | 0.013 | 0.958 | 0.043 | 0.014 | 1.000 | 0.058 | 0.010 |
| Unemployment Rate | 0.951 | -0.062 | 0.022 | 0.958 | -0.064 | 0.022 | - | - | - |
| Labor Productivity | 0.943 | 0.011 | 0.004 | 0.907 | 0.011 | 0.005 | 0.958 | 0.012 | 0.004 |
| GDP per Capita | 0.718 | -0.010 | 0.007 | 0.662 | -0.009 | 0.008 | - | - | - |
| Wage | 0.678 | -0.010 | 0.008 | 0.687 | -0.010 | 0.008 | 0.998 | -0.021 | 0.005 |
| GVA Industry | - | - | - | - | - | - | 0.997 | 0.015 | 0.010 |
| *CEE/Candidates - Dummy interactions* | | | | | | | | | |
| Candidates | | | | - | - | - | 1.000 | 0.064 | 0.010 |
| Candidates×Population with Tertiary Education | | | | | | | 1.000 | -0.203 | 0.028 |
| Candidates×GVA Industry | | | | | | | 0.997 | -0.099 | 0.018 |
| Share of posterior probabilities - Best model | | 0.06 | | | 0.05 | | | 0.17 | |
| Share of posterior probabilities - Top 25 models | | 0.43 | | | 0.36 | | | 0.65 | |
| Share of posterior probabilities - Top 50 models | | 0.51 | | | 0.43 | | | 0.73 | |
| **Corr PMP** | | 0.9999 | | | 1.0000 | | | 0.9999 | |

# Between Countries Baseline Model

| | *Model 1* | | | *Model 2* | | | *Model 3* | | |
|---|---|---|---|---|---|---|---|---|---|
| **Variable** | **PIP** | **PM** | **PSD** | **PIP** | **PM** | **PSD** | **PIP** | **PM** | **PSD** |
| GVA Public | 0.998 | -1.008 | 0.207 | - | - | - | - | - | - |
| Life Expectancy | 0.974 | -0.030 | 0.007 | - | - | - | - | - | - |
| Fertility Rate | 0.790 | 0.091 | 0.055 | - | - | - | - | - | - |
| Population with Tertiary Education | - | - | - | 0.944 | 0.473 | 0.230 | - | - | - |
| Migration Rate | - | - | - | 0.924 | -5.300 | 2.095 | - | - | - |
| GVA Industry | - | - | - | 0.690 | 0.273 | 0.207 | 0.761 | 0.342 | 0.220 |
| GVA Agriculture | - | - | - | - | - | - | 0.820 | -1.472 | 0.821 |
| *CEE/Candidates - Dummy interactions* | | | | | | | | | |
| Candidates | | | | 1.000 | 0.459 | 0.048 | 1.000 | 0.268 | 0.153 |
| CEE | | | | 1.000 | 0.255 | 0.040 | 1.000 | 0.266 | 0.120 |
| Candidates×GVA_agriculture | | | | | | | 0.819 | 2.276 | 1.182 |
| Share of posterior probabilities - Best model | | 0.19 | | | 0.31 | | | 0.22 | |
| Share of posterior probabilities - Top 25 models | | 0.68 | | | 0.74 | | | 0.77 | |
| Share of posterior probabilities - Top 50 models | | 0.75 | | | 0.82 | | | 0.83 | |
| **Corr PMP** | | 0.9999 | | | 1.0000 | | | 1.0000 | |

*Notes:* Models with 301 regions, with GDP per capita growth rate from 2009–2019 as the dependent variable.

# Within Countries Baseline Model

| | *Model 1* | | | *Model 2* | | |
|---|---|---|---|---|---|---|
| **Variable** | **PIP** | **PM** | **PSD** | **PIP** | **PM** | **PSD** |
| Population with Basic Education | 0.989 | -0.320 | 0.078 | 0.720 | -0.218 | 0.151 |
| Wage | 0.929 | -0.121 | 0.037 | 0.818 | -0.100 | 0.052 |
| Investment Rate | 0.228 | 0.093 | 0.189 | - | - | - |
| Fixed Capital Formation Rate | 0.171 | 0.073 | 0.185 | - | - | - |
| *Dummy interactions* | | | | | | |
| CEE×GVA Agriculture | | | | 0.305 | -0.666 | 1.070 |
| Share of posterior probabilities - Best model | | 0.17 | | | 0.11 | |
| Share of posterior probabilities - Top 25 models | | 0.64 | | | 0.40 | |
| Share of posterior probabilities - Top 50 models | | 0.71 | | | 0.48 | |
| **Corr PMP** | | 1.0000 | | | 0.9998 | |

*Notes:* Models with 301 regions, with GDP per capita growth rate from 2009–2019 as the dependent variable. Unlike previous tables also variables with PIP $< 0.5$ are shown because the model results are very constrained. Countries fixed effects are **not** included as fixed regressors because otherwise no other variable would result significant.

# SAR Baseline Model

| | *Model 1* | | | *Model 2* | | | *Model 3* | | |
|---|---|---|---|---|---|---|---|---|---|
| **Variable** | **PIP** | **PM** | **PSD** | **PIP** | **PM** | **PSD** | **PIP** | **PM** | **PSD** |
| Wage | 0.719 | -0.077 | 0.056 | 0.661 | -0.070 | 0.057 | 0.831 | -0.088 | 0.045 |
| Migration Rate | 0.424 | -2.165 | 2.731 | 0.404 | -2.065 | 2.706 | 0.135 | -0.689 | 1.863 |
| GVA Public | 0.407 | -0.248 | 0.322 | 0.333 | -0.199 | 0.303 | 0.142 | -0.086 | 0.223 |
| Share of posterior probabilities - Best model | | 0.31 | | | 0.32 | | | 0.63 | |
| Share of posterior probabilities - Top 25 models | | 0.71 | | | 0.71 | | | 0.91 | |
| Share of posterior probabilities - Top 50 models | | 0.78 | | | 0.78 | | | 0.94 | |
| **Corr PMP** | | 1.0000 | | | 1.0000 | | | 1.0000 | |

*Notes:* Models with 301 regions, with GDP per capita growth rate from 2009–2019 as the dependent variable. Unlike previous tables also variables with PIP $< 0.5$ are shown because the model results are very constrained.

# Robustness Check

**Robustness checks:**

- Imputation method (`mice`)
- Multicollinearity (`mprior = dilut`);
- *Distance decay parameter alternatives*;
  - inverse distance matrix with $\phi = 1$ the only dominant weight matrix, with PMP = 100%.
- *Unconstrained Durbin Model.*

# *Q&A*

Point of discussion:

- Dataset of choice (pool/cross),
- Questions on `spatFiltering()`: repetition eigenvalues across years,
- How to implement Durbin Model structure in `spatBMS`,
- How to get different estimates for different Spatial Matrices,
- Correlation Coefficient for overall spatial model.
  - ▷ significant correlation is only observed in the cross dataset, not the pool → could be because of repetition same regions across years.

All resources are available on: GitHub

# *Q&A*

Point of discussion:

- Dataset of choice (pool/cross),
- Questions on `spatFiltering()`: repetition eigenvalues across years,
- How to implement Durbin Model structure in `spatBMS`,
- How to get different estimates for different Spatial Matrices,
- Correlation Coefficient for overall spatial model.
  - ▷ significant correlation is only observed in the cross dataset, not the pool → could be because of repetition same regions across years.

All resources are available on: GitHub

**Thank you for your time!**

# References I


Jesús Crespo Cuaresma, Gernot Doppelhofer, and Martin Feldkircher. The determinants of economic growth in european regions. Regional Studies, 48(1):44–67, 2014.

Martin Feldkircher and Stefan Zeugner. Bayesian model averaging employing fixed and flexible priors: The bms package for r. Journal of Statistical Software, 68, 2015. doi: 10.18637/jss.v068.i04.

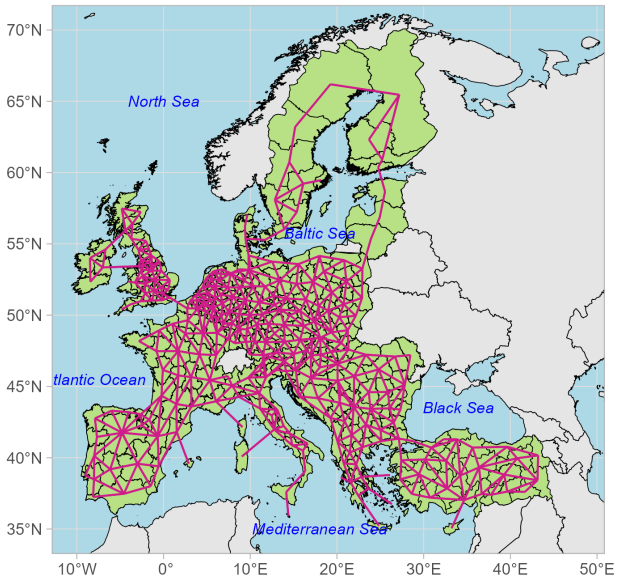
Carmen Fernandez, Eduardo Ley, and Mark FJ Steel. Benchmark priors for bayesian model averaging. Journal of Econometrics, 100 (2):381–427, 2001.

Robert E Kass and Adrian E Raftery. Bayes factors. Journal of the american statistical association, 90(430):773–795, 1995.

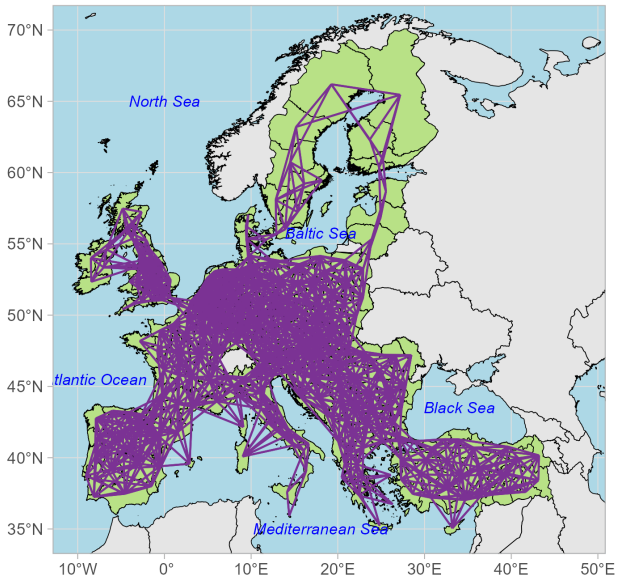
# Spatial Matrices

**Comparison of Spatial Connections**

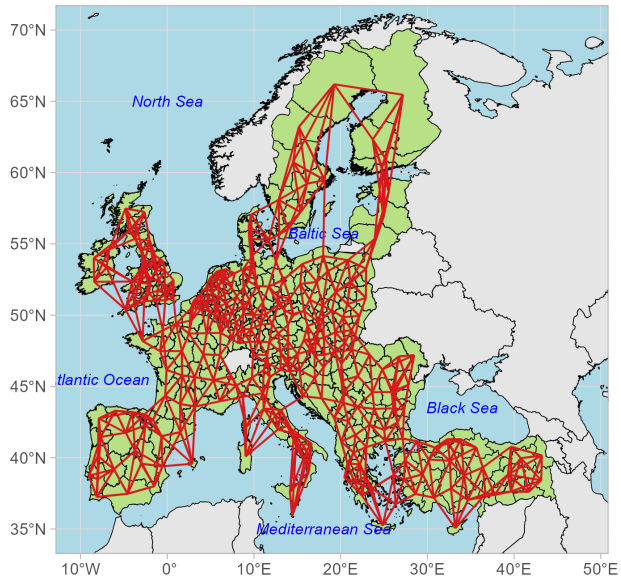
**Fist Order Queen Contiguity**




**Second Order Queen Contiguity**

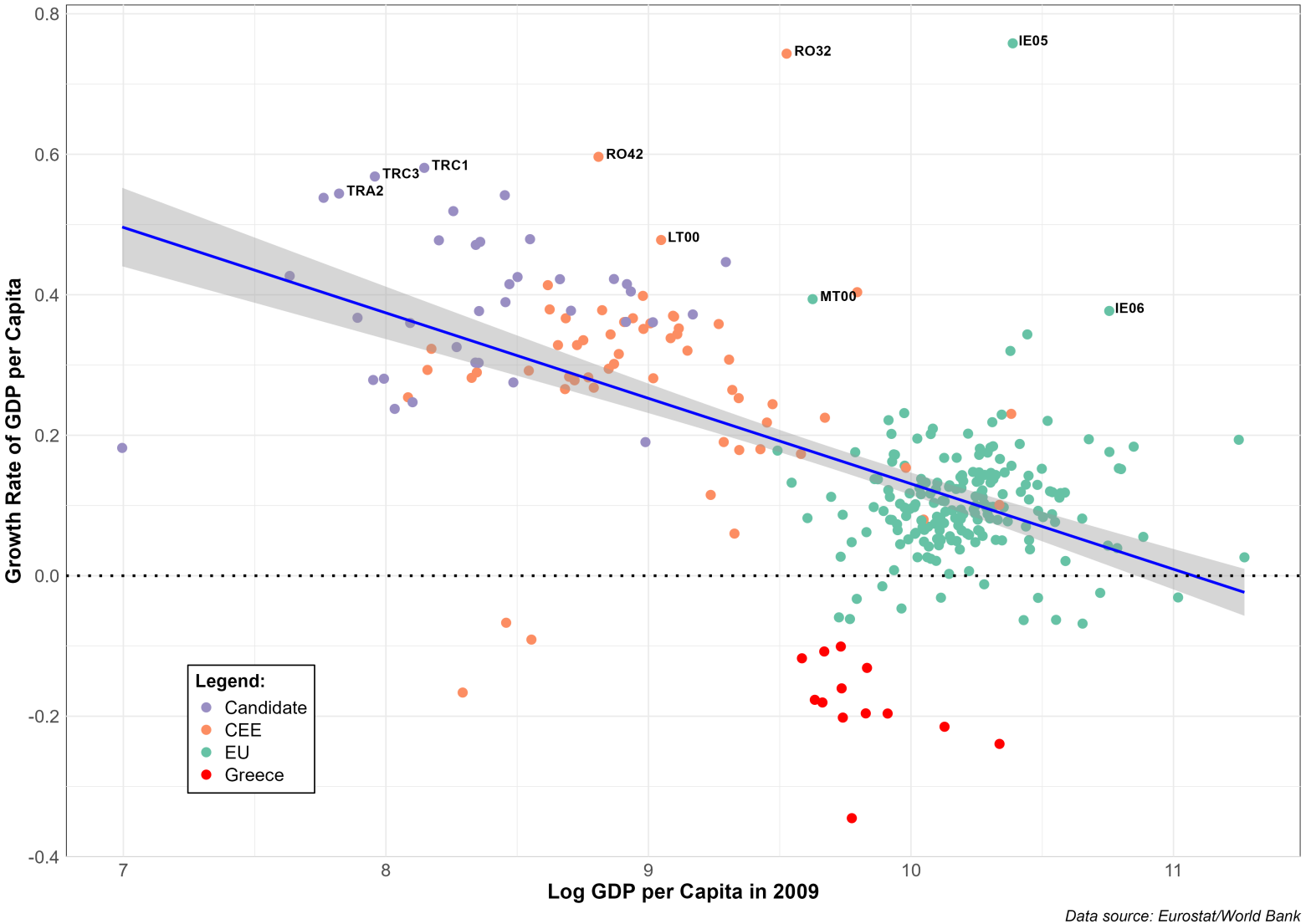



**K Neighbourghs**




Source: Eurostat, GADM

# Beta Convergence




*Data source: Eurostat/World Bank*