Master Thesis Conference

Regional Growth Determinants across the European Union and its Candidates

Elia Di Gregorio

Thesis Supervisor: Prof. Jesús Crespo Cuaresma

Vienna University of Economics and Business

Department of Economics

25 October 2024

Table of Contents

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). Results:
  - **Income Convergence:** Strong convergence in newer EU member states, especially in **capital** regions;
  - **Human Capital:** Higher education share boosts GDP growth;
  - **Spatial Spillovers:** Positive but limited spillover effects from explanatory variables.

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). Results:
  - **Income Convergence:** Strong convergence in newer EU member states, especially in **capital** regions;
  - **Human Capital:** Higher education share boosts GDP growth;
  - **Spatial Spillovers:** Positive but limited spillover effects from explanatory variables.

Research aligns with larger literature on:

Regional Growth Determinants and BMA	Convergence and Regional Disparities	Policy Implications and Cohesion

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.

Dataset

This study covers **265 EU** and **36 Candidate** regions based on **NUTS-2** classification and counts **30 regressors**, with *growth of regional GDP per capita* as dependent variable.

Thematic Groups:

1. Factor accumulation and convergence;
2. Demography;
3. Human capital;
4. Sectoral structure and employment;
5. Socio-geographical;
6. Interaction terms.

Data Sources:

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

More Info

Basian Model Averaging

- BMA addresses model uncertainty by averaging across models, rather than selecting a single "best" model → models are weighted by their ability to explain the data.
  - ▷ This prevents issues like overfitting and multicollinearity.

- Incremental model specifications:
  1. **Cross-sectional baseline** model for regions;
  2. Baseline model expanded with **country fixed effects** model to control for unobserved heterogeneity;
  3. Baseline adjusted by **Spatial Autoregressive** (SAR) model to account for 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:

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

- γ is an N-dimensional column vector of stacked growth rates of income per capita for N regions;
- ι_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;
- ρ is a scalar indicating the degree of spatial autocorrelation;
- ϵ is an error term which may contain country-specific fixed effects.

Bayes' Theorem

BMA averages over 2^K models, M_j $(j = 1, ..., K)$, with weights based on posterior model probabilities (PMPs).

- PMP is derived using Bayes' Theorem:

$$p(M_j|\mathcal{D}) = \frac{p(\mathcal{D}|M_j)p(M_j)}{\sum_{j=1}^{2^K} p(\mathcal{D}|M_j)p(M_j)}$$

  where $p(M_j)$ is the prior on model M_j and $p(\mathcal{D}|M_j)$ is the likelihood of M_j.

- The resulting posterior distribution of coefficient β_k is:

$$p(\beta_k|\mathcal{D}) = \sum_{j=1}^{2^K} p(\beta_k|M_j, \mathcal{D})p(M_j|\mathcal{D})$$

Parametrization

- **Parameter Prior:** Zellner's g-prior balances between the number of observations and the number of variables, following a standard benchmark approach.
- **Model Prior:** A binomial-beta prior that sets an expected model size of half the total number of variables.
- **MCMC Sampling (MC^3):** Efficiently searches the model space without evaluating all possible models.
  - The Birth-Death sample was used;
  - 10 million draws and a 3 million iteration burn-in period;
  - Strong-heredity Principle.

More Info

Spatial Weight Matrices in SAR Model

- The Spatial Autoregressive (SAR) model accounts for spatial dependencies by using multiple spatial weight matrices:

  1. **Inverse Distance:** Weights decrease with distance, using different rates of decline;
  2. **Queen-Contiguity:** First-order (direct neighbors) and second-order (neighbors-of-neighbors);
  3. **k-Nearest Neighbors (k-NN):** Defines spatial relationships by the $k = 5$ closest neighbors.

To Figure

- Combining different matrices helps ensure robustness by capturing varied spatial structures.

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	0.074	0.013	0.813	0.042	0.025	0.717	0.029	0.022
GVA Construction	0.999	-0.150	0.031	0.999	-0.147	0.034	0.998	-0.147	0.032
Investment Rate	0.999	0.052	0.010	-	-	-	-	-	-
GDP per Capita	0.998	-0.017	0.005	-	-	-	-	-	-
GVA Public	0.997	-0.059	0.014	-	-	-	-	-	-
Pop. with Tertiary Edu.	0.986	0.055	0.010	0.983	0.050	0.011	1.000	0.046	0.009
Migration Rate	0.846	-0.273	0.143	0.991	-0.385	0.096	-	-	-
Labor Productivity	0.535	0.005	0.005	0.520	0.006	0.007	0.999	0.018	0.004
GVA Industry	-	-	-	0.690	0.273	0.207	1.000	0.025	0.012
Capital City	-	-	-	0.795	-0.006	0.004	1.000	-0.002	0.003
Wage EUR	-	-	-	-	-	-	0.996	-0.020	0.005
Distance from Buxelles	-	-	-	-	-	-	0.980	-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	-0.009	0.010
Candidates × GVA Industry							1.000	-0.153	0.020
CEE × Pop. Tertiary Edu							1.000	0.142	0.023
Candidates × Pop. with Tertiary Edu.							1.000	-0.194	0.031
Candidates × Capital City							0.983	-0.025	0.007
CEE × Capital City							0.999	-0.026	0.005
Share of posterior probabilities - Best model		0.25			0.14			0.11	
Share of posterior probabilities - Top 25 models		0.80			0.71			0.45	
Share of posterior probabilities - Top 50 models		0.87			0.78			0.53	
Corr PMP		1.0000			1.0000			0.9999	
Adjusted R^2		0.37			0.39			0.42	

Notes: PIP, posterior inclusion probability; PM, posterior mean; PSD, posterior standard deviation. Time fixed effects were included across all model specifications. Model 1: Cross-section of regions (baseline). Model 2: Cross-section of regions including the Central and Eastern Europe (CEE) and Candidates dummy variable. Model 3: Cross-section of regions further including interaction terms and strong heredity principle application.

Within Countries Baseline Model

	Model 1			*Model 2*		
Variable	**PIP**	**PM**	**PSD**	**PIP**	**PM**	**PSD**
GVA Construction	0.999	-0.284	0.042	0.999	-0.295	0.036
NEET Share	0.996	0.102	0.022	-	-	-
Pop. with Primary Edu.	0.900	-0.066	0.026	0.977	-0.067	0.015
Unemployment Rate	0.867	-0.069	0.033	-	-	-
GDP per Capita	0.793	-0.010	0.006	0.996	-0.025	0.007
GVA Public	-	-	-	0.650	-0.042	0.035
CEE/Candidates - Dummy interactions						
Candidates×GVA Services				0.995	0.135	0.026
CEE×GDP per Capita				0.998	0.049	0.014
Candidates×GDP per Capita				0.993	-0.041	0.008
CEE×Capital City				0.846	-0.022	0.011
CEE×GVA Public				0.665	0.142	0.111
Share of posterior probabilities - Best model		0.12			0.09	
Share of posterior probabilities - Top 25 models		0.72			0.66	
Share of posterior probabilities - Top 50 models		0.80			0.72	
Corr PMP		0.9998			1.0000	
Adjusted R^2		0.46			0.48	

Notes: PIP, posterior inclusion probability; PM, posterior mean; PSD, posterior standard deviation. Model 1: Cross-section of regions with country-fixed effects. Model 2: Cross-section of regions further including the interaction terms. Time and country fixed effects were included as fixed regressors in both models. The CEE and Candidates dummy variables were left out because of perfect multicollinearity, thus the 'strong heredity prior' was not implemented.

Spatial BMA

	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	-0.227	0.034	1.000	-0.205	0.032
NEET Share	0.994	0.081	0.020	0.994	0.082	0.020	0.719	0.029	0.022
Pop. with Tertiary Edu.	0.968	0.044	0.013	0.947	0.043	0.015	1.000	0.058	0.011
Unemployment Rate	0.954	-0.062	0.022	0.960	-0.064	0.022	-	-	-
Labor Productivity	0.950	0.012	0.004	0.911	0.011	0.005	0.960	0.012	0.004
GDP per Capita	0.719	-0.010	0.007	0.661	-0.009	0.008	-	-	-
Gross Wage	0.694	-0.010	0.008	0.698	-0.010	0.007	0.998	-0.021	0.005
GVA Industry	-	-	-	-	-	-	0.996	0.015	0.010
CEE/Candidates - Dummy interactions									
Candidates							1.000	0.064	0.011
Candidates×Pop. with Tertiary Edu.							1.000	-0.203	0.028
Candidates×GVA Industry							0.996	-0.099	0.018
Share of posterior probabilities - Best model		0.07			0.05			0.17	
Share of posterior probabilities - Top 25 models		0.46			0.37			0.65	
Share of posterior probabilities - Top 50 models		0.54			0.45			0.74	
Corr PMP		0.9999			0.9998			0.9948	
Adjusted R^2		0.37			0.39			0.42	

Notes: PIP, posterior inclusion probability; PM, posterior mean; PSD, posterior standard deviation. Model 1: Cross-section of regions with spatial eigenvectors. Model 2: Cross-section of regions including the Central and Eastern Europe (CEE) and Candidates variable. Model 3: Cross-section of regions further including interaction terms. Under Model 3 the 'strong heredity prior' was employed. Time fixed effects were not included as fixed regressors because of the `spatFilt.bms()` function argument's limitation.

Robustness Check

1. Imputation method (`amelia vs. mice`);
2. Multicollinearity (`mprior = dilut`);
3. Combination of Spatial Weight Matrices; To Figure
4. Unconstrained Durbin Model. To Table

Key Findings

- **Human capital** is a strong driver of regional growth across Europe.
- **Sectoral composition** is crucial:
  - *Industrial* activity boosts growth;
  - Over-reliance on *construction* hinders long-term productivity.
- Evidence of **convergence** between EU and Candidate countries and of **spatial spillovers**.
- **Policy implications**:
  - Invest in education and skills;
  - Promote industrial development and diversify economic structures;
  - Enhance interregional knowledge exchange;
  - Tailored strategies needed for Candidate countries to address structural challenges;

References

All resources are available on: GitHub.

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

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

Back to Overview

Appendix: NUTS-2 Specification

NUTS Regions are defined based on Population Thresholds:

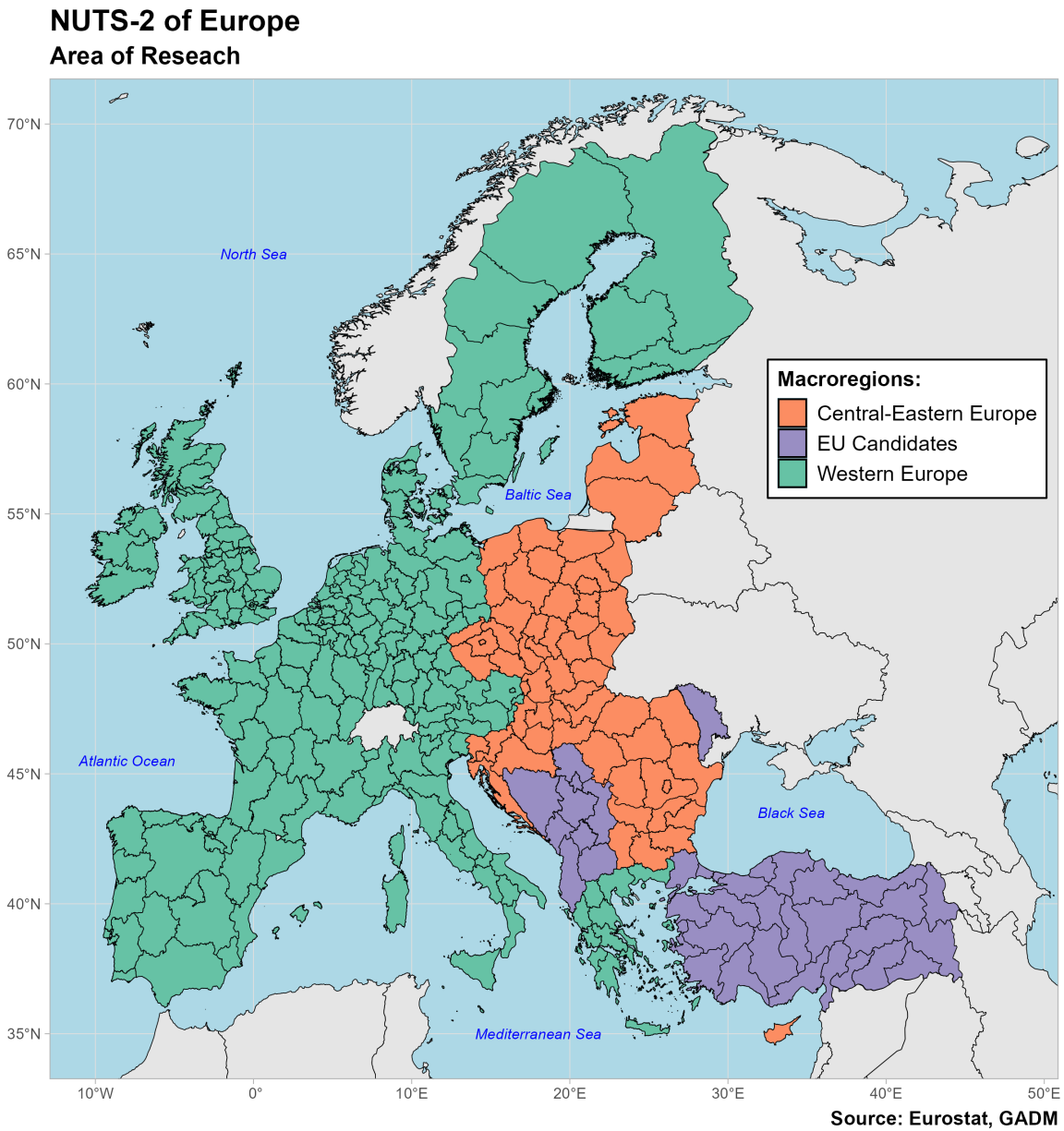
NUTS Level	Min. Pop	Max. Pop
NUTS-1	3,000,000	7,000,000
NUTS-2	800,000	3,000,000
NUTS-3	150,000	800,000

Not all Candidate countries follow NUTS system, and/or have good regional data availability. Overall:

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

In some cases, EU territorial units diverge from the 2021 NUTS Classification System to ensure consistent and comprehensive data coverage over time across EU.

Appendix: Regional Coverage



Appendix: 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 EUR09 (log);
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. Sectoral components of GVA;
14. Output density;
15. Employment density (km^2);
16. Population density (km^2);
17. Distance from Bruxelles (km);
18. Dummy variables and interactions.

Appendix: Variables Breakdown

Sectoral Components of GVA[1]:

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

Educational Attainment[2]**:**

- 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×

- Capital city region;
- Population with tertiary education;
- Sectoral components of GVA.

Back

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

Appendix: Prior Structure

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 $K/2$.

Appendix: Markov Chain Monte Carlo

Birth-death sampler used to add or drop covariates systematically:

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

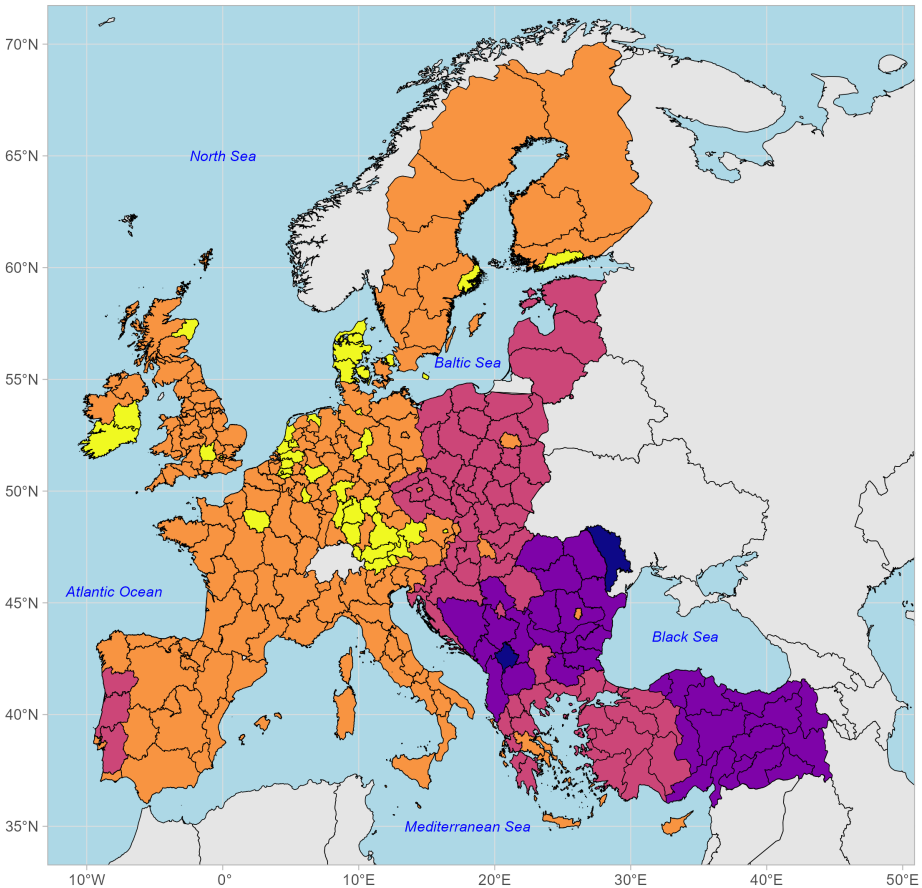
1. This sampler starts at model M_i with a posterior model probability of $p(M_i|\mathcal{D})$;
2. One of the k potential covariates in X_k is randomly chosen;
3. If the covariate is part of M_i, it is dropped ("dies"), otherwise it is added ("born") to create another model M_j;
4. The sampler now switches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

Strong Heredity Principle: interaction terms are only included in models when their corresponding main effects (i.e., the individual interacted variables) are also present.

Appendix: Distribution Effects

Spatial Distribution PMs

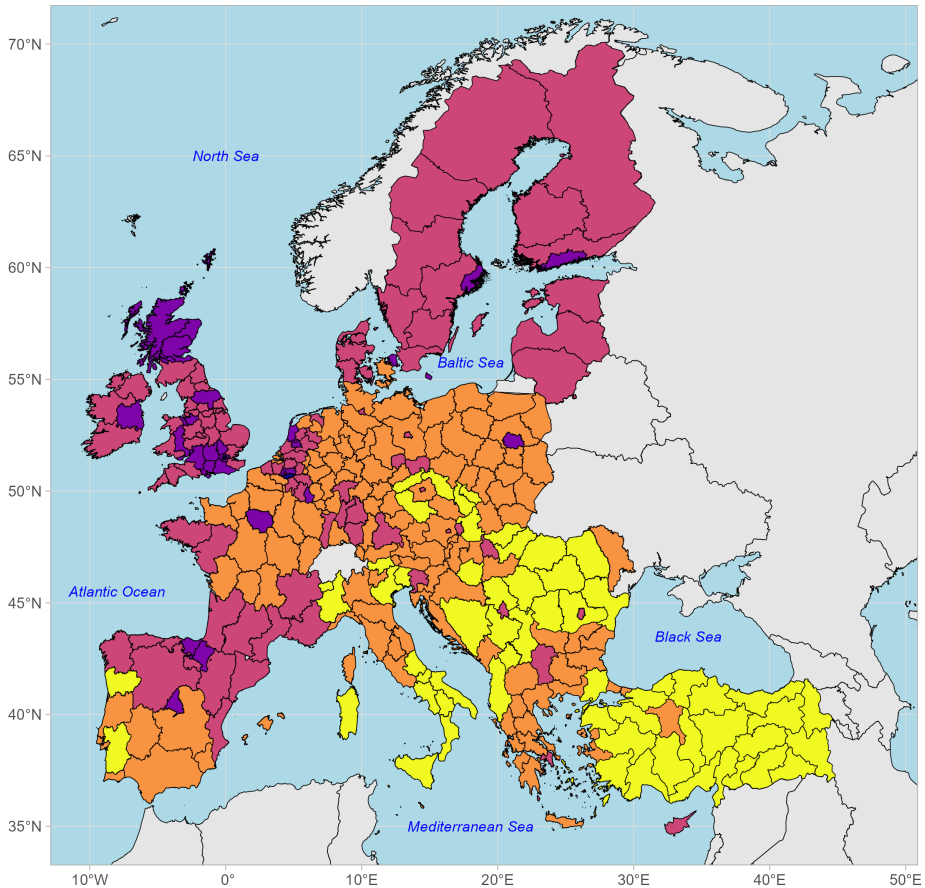
Initial Income



Average estimated effect:

- -2.03% to -1.88%
- -1.88% to -1.72%
- -1.72% to -1.57%
- -1.57% to -1.42%
- -1.42% to -1.26%

Human Capital



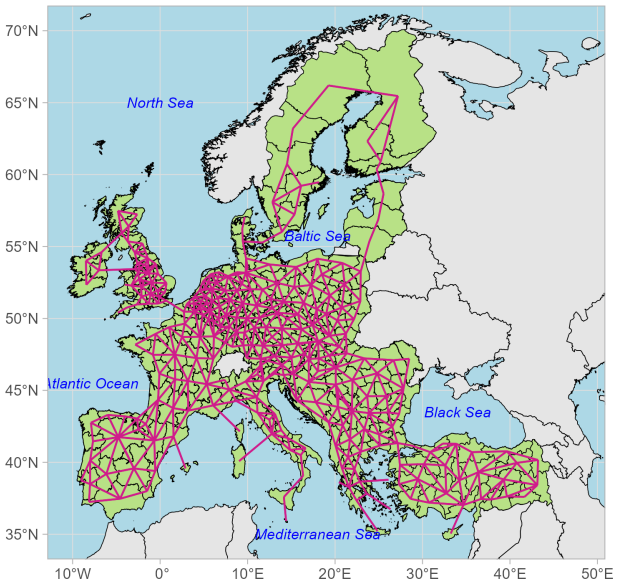
Average estimated effect:

- 0.32% to 0.83%
- 0.83% to 1.34%
- 1.34% to 1.86%
- 1.86% to 2.37%
- 2.37% to 2.88%

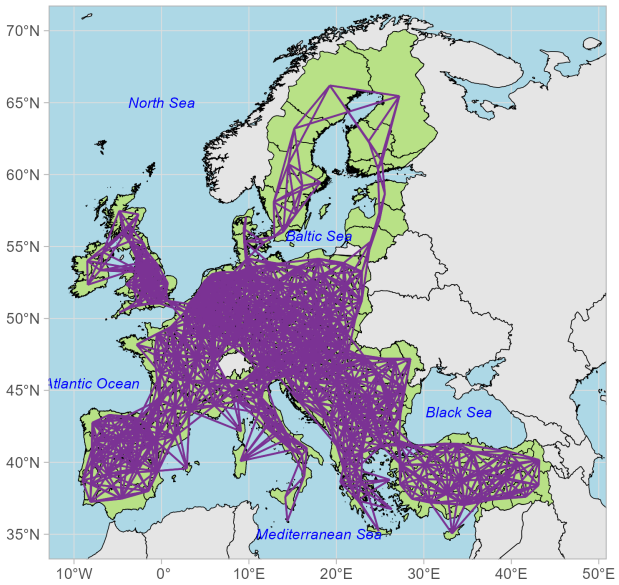
Appendix: Spatial Networks

Comparison of Spatial Networks

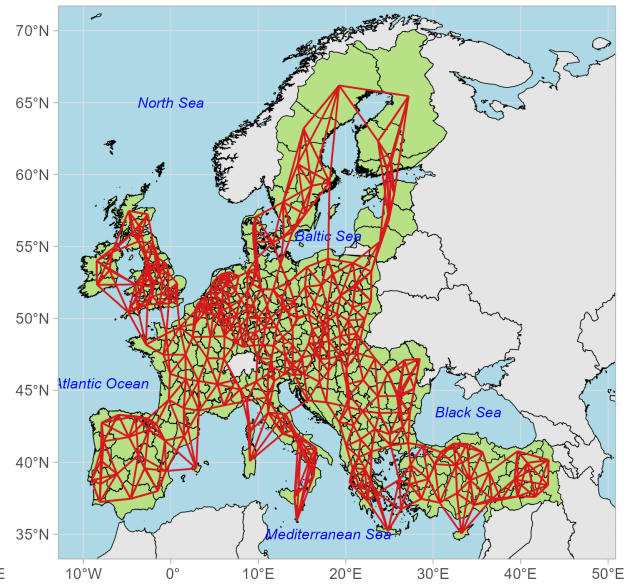
Fist Order Queen Contiguity



Second Order Queen Contiguity



K Neighbourghs



Source: Eurostat, GADM

Back

Appendix: Spatial Weight Matrices Impact

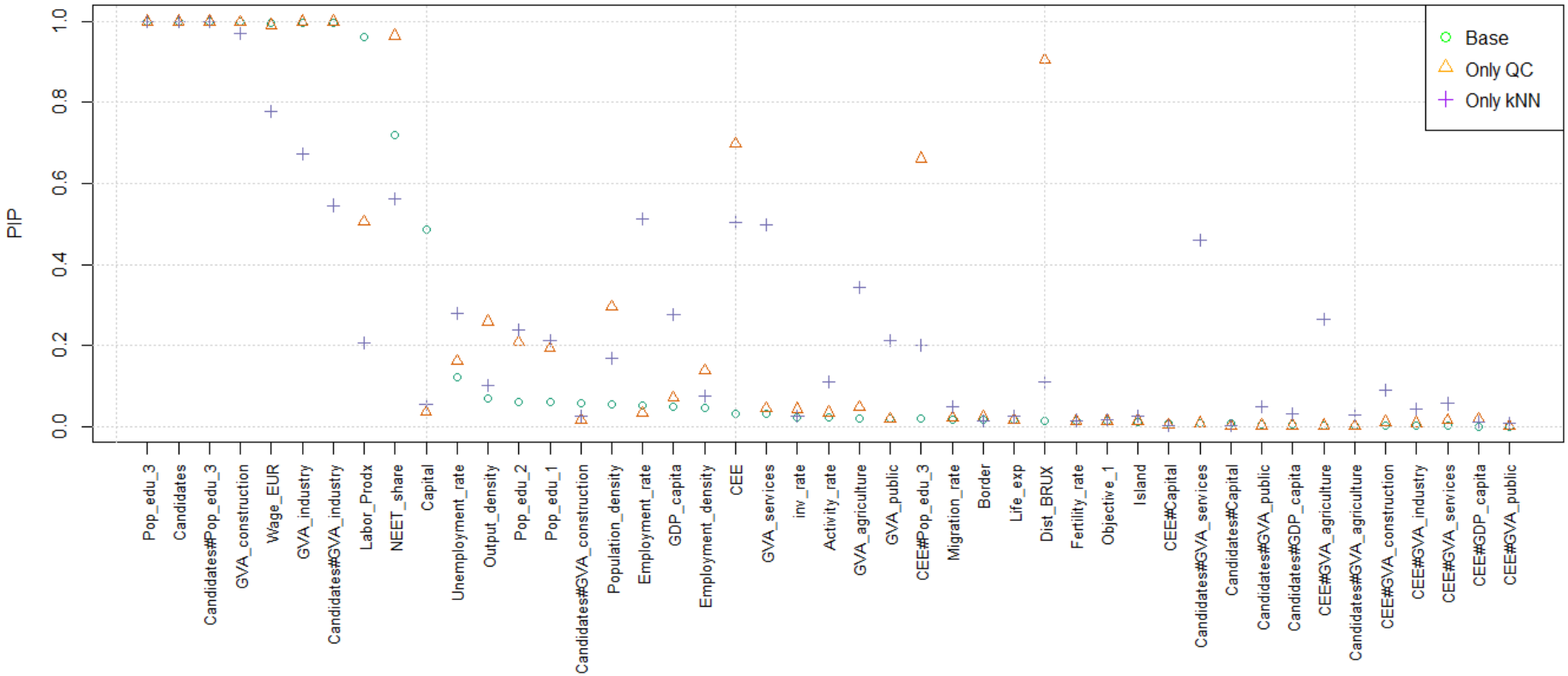


Figure: Effects of Spatial Matrices on PIPs

Effect of PMs, the model coefficients, is instead negligible.

Appendix: Unrestricted Spatial Durbin Model

	Model 1			*Model 2*			*Model 3*		
Variable	**PIP**	**PM**	**PSD**	**PIP**	**PM**	**PSD**	**PIP**	**PM**	**PSD**
W GDP per Capita	1.000	-0.0443	0.0092	0.999	-0.0255	0.0056	0.999	-0.0242	0.0052
W GVA Construction	1.000	-0.347	0.0448	1.000	-0.318	0.0438	1.000	-0.367	0.0404
W Labor Productivity	0.999	0.0264	0.0063	0.770	0.0171	0.0110	1.000	0.0382	0.0057
W Unemployment Rate	0.998	-0.112	0.0177	-	-	-	-	-	-
Wage	0.758	-0.0109	0.0067	-	-	-	0.999	-0.0213	0.0067
Pop. with Tertiary Edu.	0.981	0.0431	0.0108	0.975	0.0432	0.0112	0.996	0.0619	0.0086
GVA Industry	-	-	-	-	-	-	0.888	0.0280	0.0130
Output Density	0.520	-0.0001	0.0001	-	-	-	0.690	-0.00004	0.00005
Capital City	-	-	-	0.909	-0.0079	0.0032	-	-	-
Migration Rate	-	-	-	0.850	-0.319	0.158	-	-	-
CEE/Candidates - Dummy interactions									
Candidates				1.000	0.0416	0.0038	0.654	0.0349	0.0288
CEE				1.000	0.0258	0.0032	-	-	-
W Candidates×Capital							1.000	-0.0480	0.0082
W CEE×Capital City							0.999	-0.0454	0.0071
Candidates×Pop. with Tertiary Edu.							0.999	-0.204	0.0427
W CEE×Pop. with Tertiary Edu.							0.993	0.148	0.0245

Notes: PIP, posterior inclusion probability; PM, posterior mean; PSD, posterior standard deviation. Model 1: Cross-section of regions with spatially lagged variables. Model 2: Cross-section of regions including CEE and Candidates variables. Model 3: Cross-section of regions with interaction terms. Spatial lags of variables are denoted by *W*.