Spatial Distribution Dynamics

S. Magrini M. Gerolimetto
Department of Economics – Università Ca' Foscari Venezia

Pisa, 22 November 2019

Motivation and outline of the presentation

Motivation

- it is quite common in convergence analyses across spatial units (countries, regions) that data exhibit **strong spatial dependence**
- neglecting spatial dependence may affect the results

⇒ develop a tool for the analysis of cross-sectional convergence within the distribution dynamics approach when data are spatially dependent

Outline

- recall some spatial dependence issues
- discuss consequences of spatial dependence on the analysis of distribution dynamics
- develop a two-step spatial nonparametric estimator for adjusting existing tools in distribution dynamics analysis
- analyze convergence among US states

Modeling spatial dependence

Time series analysis $\rightarrow Y_t = \beta_1 + \beta_2 X_{2t} + u_t, \; t = 1, ..., T$

Spatial analysis $\rightarrow Y_i = \beta_1 + \beta_2 X_{2i} + u_i, \; i = 1, ..., N$, where some spatial interaction effects can be included.

- Endogenous interaction effects (spatial lag of Y)
- Exogenous interaction effects (spatial lag of X)
- Interaction effects among error terms (spatial dependence in u)

It is possible to think of general nesting spatial models

$$Y = \rho WY + X\beta + WX\gamma + u, \quad u = \lambda Wu + \epsilon$$

- SAR: $\gamma = 0, = 0, \Rightarrow Y = \rho WY + X\beta + u$
- SLX: $\rho = 0, = 0, \Rightarrow Y = X\beta + WX\gamma + u$
- SEM: $\gamma = 0, \rho = 0, \Rightarrow Y = X\beta + u$, where $u = \lambda Wu + \epsilon$

The structure of spatial interactions

It is necessary to impose a structure on the extent of spatial interaction.

One possible approach - following a neighborhood view - is to define a neighborhood set $N(i)$ for each location i.

By doing this, it is possible to specify for the neighbor set a

- a contiguity form
- or a distance decay based form
  - distance band
  - k-nearest neighbors
  - more complex distance decay functions (e.g. inverse of distance)

The outcome is a **spatial weights matrix**.

The spatial weight matrix

The spatial weight matrix W is $N \times N$ positive matrix with elements w_{ij} and it represents, for each location i in the system, which of the other locations in the system affect i.

In its simplest version, the W is in binary contiguity form, where $w_{ij} = 1$ for i and j neighbors (e.g. $d_{ij} <$ critical distance), $w_{ij} = 0$ otherwise $w_{ii} = 0$ by convention.

More in general, weights can be defined according to:

- Contiguity
  - common boundary (regularly or irregularly located units)
- Distance
  - distance band
  - k-nearest neighbors
- Other (even more general)
  - social distance
  - complex distance decay functions

The W matrix as spatial shift operator

In contrast to the unambiguous concept of time shift along the time axis, **there is no such a corresponding concept in space**, especially when observations are irregularly located.

So in space we adopt the spatial weight matrix, the acts in the sense of calculating the weighted average of random variables at neighboring locations.

This is given by reading by row i the W matrix that gives the number of nonzero weighted j locations that gives

$$[Wy]_i = \sum_{j=1}^{N} w_{ij} Y_j$$

For easy of interpretation often the elements are *row standardized*, so for each i $\sum_{j=1}^{N} w_{ij}=1$. Hence it is more visible the interpretation of the spatial lag as weighted average of the neighbors, or spatial smoother.

Spatial autocorrelation specification

Together with the parameters in the models, the W matrix plays a role in the specification of the spatial autocorrelation.

There is also *another* approach to the specification of spatial autocorrelation, which is the so-called direct representation of the spatial autocorrelation.

The objective of this second approach is to express the element of the covariance matrix in a parsimonious fashion, as a direct function of the distance between locations i and j.

$$Cov(u_i, u_j) = \sigma^2 f(d_{ij}, \phi)$$

d_{ij} is the distance between sites i,j
$f(.)$ is a decaying function such that $\frac{\partial f}{\partial d_{ij}} < 0$, $|f(d_{ij}, \phi| \leq 1$
ϕ is an appropriate vector of parameters.

Distribution dynamics

The distribution dynamics approach in short

- let $F(Y_t)$ and $F(Y_{t+s})$ represent the cross-sectional distributions of per capita income at time t and $t+s$
- assume they admit a density ($f(Y_t)$ and $f(Y_{t+s})$ respectively)
- assuming the dynamics between time t and $t+s$ can be modelled as a first order process, then

$$f\left(Y_{t+s}\right) = \int_{-\infty}^{\infty} f\left(Y_{t+s}|Y_t\right)\ f\left(Y_t\right)\ dY_t$$

- convergence is analysed through:
  - an estimate of the **conditional density** (or stochastic kernel) $f\left(Y_{t+s}|Y_t\right)$, traditionally obtained via the **kernel estimator**
  - an estimate of the **ergodic** (or stationary) **distribution** (as $s \to \infty$), under the assumption that the process is Markov and time homogeneous

Conditional density estimation: kernel estimator

- the corner-stone of the approach is the **conditional density** $f\left(Y_{t+s}|Y_t\right)$
- given a sample $(Y_{1,t}, Y_{1,t+s})$, ... $(Y_{j,t}, Y_{j,t+s})$, ... $(Y_{n,t}, Y_{n,t+s})$, the most common estimator of a conditional density is the kernel estimator:

$$\hat{f}(Y_{t+s}|Y_t) = \sum_{j=1}^{n} w_j(Y_t) K_b\left(Y_{t+s} - Y_{j,t+s}\right)$$

where

$$w_j(Y_t) = \frac{K_a\left(Y_t - Y_{j,t}\right)}{\sum_{j=1}^{n} K_a\left(Y_t - Y_{j,t}\right)}$$

a, b are bandwidths controlling the degree of smoothness

Conditional density estimation: *mean-bias* issue

The mean of the conditional density $f\left(Y_{t+s}|Y_t\right)$ is the **mean function**, $M(Y_t)$

Hyndman *et al.* (1996)

- the mean function estimator implicit in the traditional kernel estimator of the conditional density is the **local constant estimator**
- the bias of the mean function estimate is carried over onto the conditional density estimate (*mean-bias*)
- the local constant estimator has poor bias properties

⇒ the local constant estimator can be replaced with other smoothers employed in nonparametric regressions $Y_{t+s} = M(Y_t) + \epsilon_t$ (*mean-bias adjustment*)

Since we are analyzing economic convergence dynamics

⇒ the mean function estimate required in the adjustment procedure is in fact an autoregression

The spatial dependence issue

Note that

- the statistical properties of $\hat{M}(Y_t)$ assume errors are zero mean and **uncorrelated**
- however, in growth and convergence studies data exhibit **spatial dependence**

$\Rightarrow$ consequences of neglecting spatial dependence in the estimate of $M(Y_t)$ are also carried over onto the conditional density estimate

Within the distribution dynamics framework

- the issue is (only rarely) tackled via spatial filtering:
  - assume that the structure of spatial dependence in known (*i.e.*, assume that the spatial weights matrix W is known)
  - filter spatial dependence away from data and then proceed with the analysis
- we follow a different route: we prefer not to make assumptions on the structure of spatial dependence

Spatial NonParametric (SNP) regression

SNP is a two-step procedure for nonparametric regression with spatially dependent data whose specific features are:

- it does not require *a priori* parametric assumptions on spatial dependence
- the information on the dependence structure is drawn from a nonparametric estimate of the spatial covariance matrix, called **spline correlogram**.

In addition:

- can be employed to estimate the mean function required in Hyndman's *mean-bias* adjustment, thus providing a way of dealing with both the *mean-bias* and the spatial dependence issues

Spline correlogram (Bjørnstad and Falk, 2001)

Along the lines of the direct representation approach, the spline correlogram is a continuous nonparametric positive semidefinite estimator of the covariance function:

- start from the sample correlation

$$\hat{\rho}_{ij} = \frac{(z_i - \bar{z})(z_j - \bar{z})}{1/n \sum_{l=1}^{n} (z_l - \bar{z})^2}$$

- take a cubic B-spline K as a smoother

$$\tilde{\rho}(d_{ij}) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} K_a(d_{ij}/h)\hat{\rho}_{ij}}{\sum_{i=1}^{n} \sum_{j=1}^{n} K_a(d_{ij}/h)}$$

- since $\tilde{\rho}$ must be not only consistent, but also positive semidefinite, use the Fourier-filter

SNP procedure

Objective: estimate $Y = M(X) + u$

Tool: SNP procedure

0. *Pilot fit*: estimate $M(X)$ with a local polynomial smoother to obtain $\hat{u} = Y - \hat{M}(X)$
1. *Nonparametric covariance matrix estimation*: use the **spline correlogram** to obtain $\hat{V}$, the estimated **spatial covariance** matrix of $\hat{u}$ (using, simply, a distance matrix)
2. *Final fit*: run the **modified regression** $Z = M(X) + \epsilon$ where
   - $Z = \hat{M}(X) + L^{-1}\hat{u}$ replaces Y
   - L is obtained through the Cholevsky decomposition of $\hat{V}$

   $\Rightarrow$ residuals ϵ are free from spatial dependence

Properties

- asymptotic properties are derived by adapting Martins-Filho and Yao's (2009) theoretical framework
- finite sample properties are established through a Monte Carlo experiment

Monte Carlo experiment

For the model

$$Y = M(X) + u$$
$$u = \rho W u + \epsilon$$

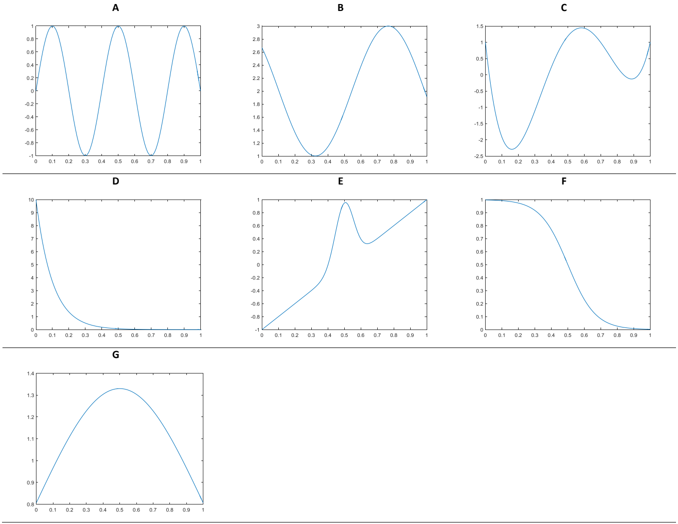
we consider a set of nonlinear functions:

A	$M(X) = sin(5\pi X)$
B	$M(X) = 2 + sin(7.1(X - 3.2))$
C	$M(X) = 1 - 48X + 218X^2 - 315X^3 + 145x^4$
D	$M(X) = 10exp(-10X)$
E	$M(X) = (-1 + 2X) + 0.95exp(-40(-1 + 2X)^2)$
F	$M(X) = 1/(1 + exp(-6 + 12X))$
G	$M(x) = (0.3\sqrt{2\pi})^{-1} exp(-2(X - 0.5)^2)$

where:

$X \sim U(0, 1)$
$\epsilon \sim N(0, \sigma^2)$, where σ is set to obtain pseudo-$R^2 = 0.2, 0.5, 0.8$
$\lambda = 0.3, 0.5, 0.8$
two W matrices (10% neighbors and contiguity from Voronoi tessellation)
sample size $(N) = 50, 100, 200$
1000 Monte Carlo replications
two types of bandwidth (direct plug-in and cross-validation minimization)

Monte Carlo functions



Monte Carlo results

Ratio (SNP over NP) of the median across replications of the MISE

		A			B			C		
pseudo-R^2	**n**	**0.3**	**0.5**	**0.8**	**0.3**	**0.5**	**0.8**	**0.3**	**0.5**	**0.8**
0.2	**50**	0.98	0.97	0.95	0.98	0.99	0.94	1.02	1.00	0.95
0.2	**100**	0.98	0.94	0.89	1.02	0.98	0.88	0.98	0.94	0.87
0.2	**200**	0.97	0.93	0.79	0.98	0.93	0.79	0.97	0.90	0.83
0.5	**50**	1.00	0.98	0.94	1.00	0.98	0.89	1.02	1.01	0.92
0.5	**100**	0.99	0.93	0.81	1.01	0.95	0.81	0.99	0.96	0.82
0.5	**200**	0.98	0.93	0.75	0.98	0.92	0.73	0.97	0.91	0.79
0.8	**50**	1.00	0.98	0.89	1.00	0.97	0.90	1.02	0.99	0.89
0.8	**100**	1.00	0.95	0.81	0.99	0.96	0.78	1.01	0.96	0.82
0.8	**200**	0.98	0.92	0.73	0.98	0.93	0.77	0.98	0.93	0.76

		D			E			F		
pseudo-R^2	**n**	**0.3**	**0.5**	**0.8**	**0.3**	**0.5**	**0.8**	**0.3**	**0.5**	**0.8**
0.2	**50**	1.02	1.00	0.94	0.99	0.97	0.92	1.01	0.96	0.95
0.2	**100**	1.01	0.95	0.88	0.98	0.96	0.88	0.99	0.97	0.84
0.2	**200**	0.98	0.96	0.85	0.99	0.96	0.79	0.98	0.94	0.79
0.5	**50**	1.01	0.98	0.94	1.00	0.97	0.90	1.02	0.98	0.97
0.5	**100**	1.02	0.95	0.88	1.00	0.95	0.89	1.01	0.97	0.87
0.5	**200**	0.99	0.92	0.78	0.99	0.92	0.86	0.99	0.94	0.82
0.8	**50**	1.01	0.98	0.94	1.00	0.98	0.92	1.00	0.98	0.98
0.8	**100**	0.99	0.97	0.82	1.00	0.97	0.84	1.01	0.96	0.89
0.8	**200**	0.99	0.93	0.74	0.98	0.92	0.77	0.99	0.94	0.78

		G		
pseudo-R^2	**n**	**0.3**	**0.5**	**0.8**
0.2	**50**	0.99	0.98	0.89
0.2	**100**	0.99	0.96	0.87
0.2	**200**	0.98	0.96	0.82
0.5	**50**	1.00	0.98	0.92
0.5	**100**	0.98	0.94	0.88
0.5	**200**	0.99	0.92	0.80
0.8	**50**	1.01	0.97	0.91
0.8	**100**	1.00	0.95	0.87
0.8	**200**	0.98	0.94	0.77

Conclusions on MC experiments

Results show

- SNP **outperforms** polynomial regression (NP)
- this is confirmed:
  - for various functional forms
  - for all considered ρ values
  - for all considered sample sizes
  - for all considered **pseudo**-R^2 values

Hence

- SNP is a valuable tool for nonparametric regression when data are spatially dependent
- SNP can be used to estimate the mean function within Hyndman's *mean-bias* adjustment thus improving the properties of the conditional density estimator

Data

USA context

- 48 coterminous US states
- quarterly data on personal per capita income (1971:Q1-2010:Q4)
- orthodromic distance between state capitals

Allow for short-run, cyclical dynamics

- the object of interest to convergence analysts is, essentially, the evolution of **potential** output
- measured output is a **noisy** indicator of potential output, contaminated by business cycle dynamics

$\Rightarrow$ as in Gerolimetto and Magrini (2014):

- extract the trend from each state's series via a Hodrick-Prescott filter
- apply the distribution dynamics approach to data on extracted trends

1971:Q1-1980:Q4

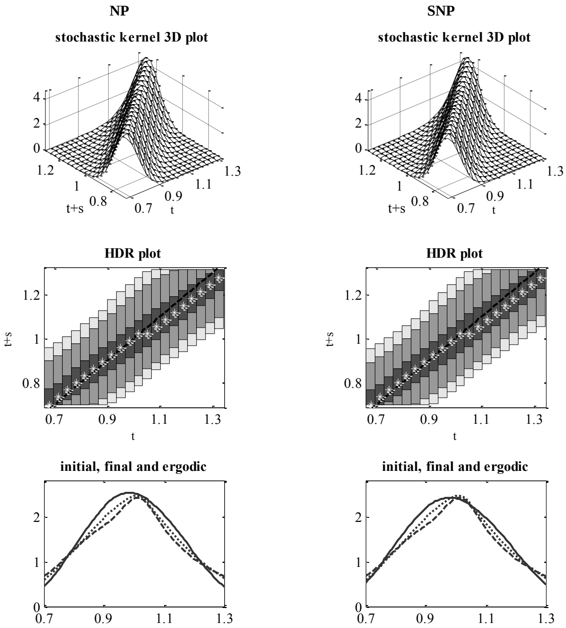


Table: Results

	Moran's I	p-value
observed initial	0.2606	0
observed final	0.1516	0.0013
filtered initial	0.2607	0
filtered final	0.1697	0.0004
residuals NP	0.0937	0.028
residuals SNP	0.0155	0.4815
	CV	IR
HP-filtered initial	0.1563	0.1905
HP-filtered final	0.1422	0.2077
ergodic NP	0.1420	0.2036
ergodic SNP	0.1463	0.2120

Table: Estimated half-life values

ergodic via SNP	ergodic via NP
3.5484	3.2585

1981:Q1-1990:Q4

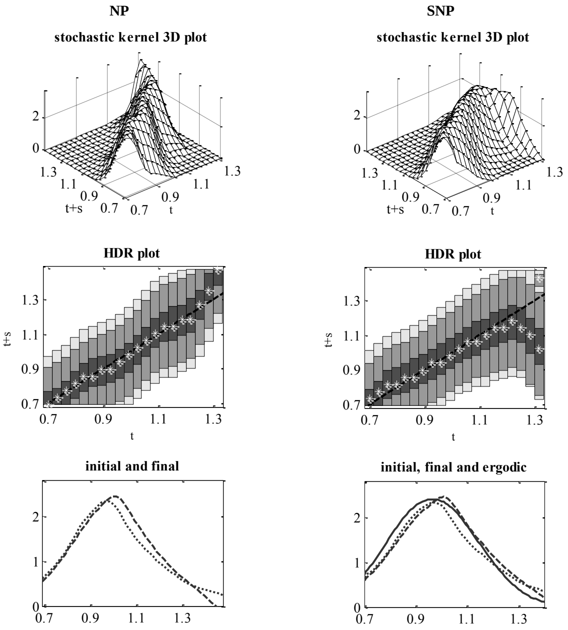


Table: Results

	Moran's I	p-value
observed initial	0.1548	0.0011
observed final	0.3014	0
HP-filtered initial	0.17	0.0004
HP-filtered final	0.3076	0
residuals NP	0.5215	0
residuals SNP	-0.0817	0.2569
	CV	IR
HP-filtered initial	0.1427	0.1905
HP-filtered final	0.1681	0.2424
ergodic NP	–	–
ergodic SNP	0.1541	0.2136

Table: Estimated half-life values

ergodic via SNP	ergodic via NP
1.7255	–

1991:Q1-2000:Q4

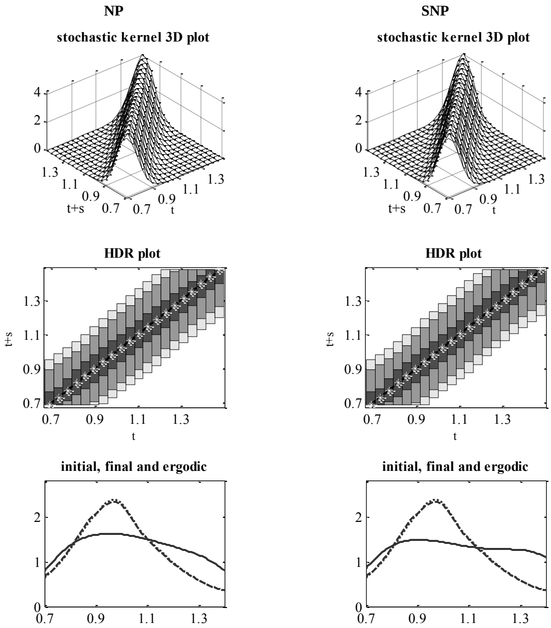


Table: Results

	Moran's I	p-value
observed initial	0.3141	0
observed final	0.2425	0
HP-filtered initial	0.306	0
HP-filtered final	0.2412	0
residuals NP	0.2007	0
residuals SNP	0.0410	0.2412
	CV	IR
HP-filtered initial	0.1677	0.2441
HP-filtered final	0.1673	0.2458
ergodic NP	0.1902	0.3158
ergodic SNP	0.2000	0.3524

Table: Estimated half-life values

ergodic via SNP	ergodic via NP
8.4571	7.0202

2001:Q1-2010:Q4

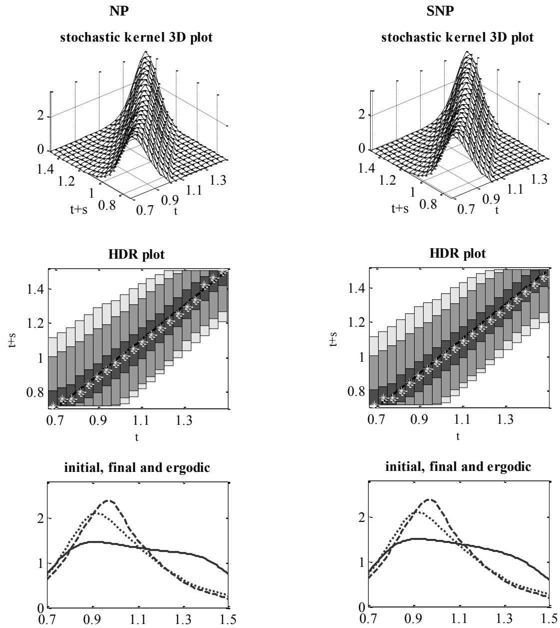


Table: Results

	Moran's I	p-value
observed initial	0.251	0
observed final	0.285	0
HP-filtered initial	0.2421	0
HP-filtered final	0.2865	0
residuals NP	0.1416	0.0019
residuals SNP	0.0042	0.6264
	CV	IR
HP-filtered initial	0.1675	0.2440
HP-filtered final	0.1803	0.2538
ergodic NP	0.2021	0.3569
ergodic SNP	0.1989	0.3420

Table: Estimated half-life values

ergodic via SNP	ergodic via NP
4.5453	4.9808

Transitional dynamics – examples

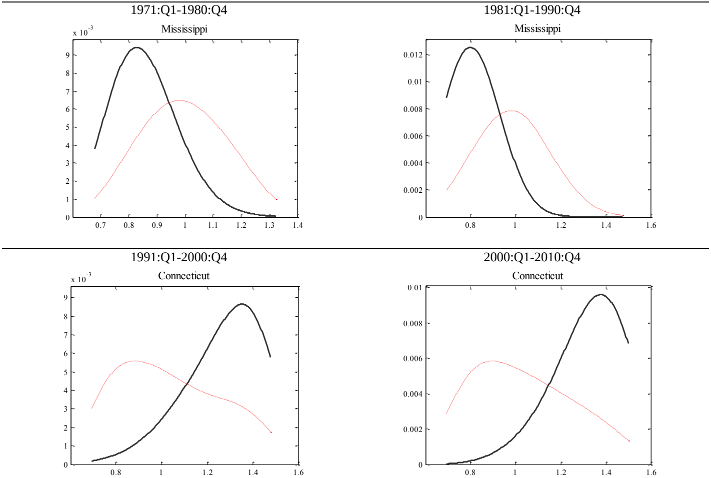


Figure: plots of the conditional density distributions for a State's initial level of (HP-filtered) per capita income after a number of iterations corresponding to the half-life and the corresponding cross-sectional distribution

Transitional dynamics

	1971:Q1-1980:Q4		1981:Q1-1990:Q4		1991:Q1-2000:Q4		2001:Q1-2010:Q4	
	3Q<1	1Q>1	3Q<1	1Q>1	3Q<1	1Q>1	3Q<1	1Q>1
Alabama	0.8512		0.8590					
Arkansas	0.8404		0.8459					
California		1.1239						
Colorado								1.2093
Connecticut		1.1482				1.3531		1.3786
Delaware		1.1266						
Illinois		1.1185						
Kentucky	0.8728		0.8656					
Louisiana	0.8647							
Maine			0.8754					
Maryland		1.1185				1.3367		1.2702
Massachusetts						1.3301		1.3244
Mississippi	0.8296		0.7999					
Nevada		1.1428						
New Jersey		1.1320				1.3465		1.3447
New Mexico		1.1239						
New York		1.1239				1.3202		
South Carolina	0.8701		0.8623					
Tennessee			0.8853					
West Virginia	0.8539		0.8492					

Table: mode of those conditional distributions (and hence those States) for which either the 3^{rd} quartile is smaller than 1 or the 1^{st} quartile is larger than 1

Conclusions

Overall, we find

- evidence of **persistence** in the 1970s and 1980s
- evidence of **club convergence** in the 1991:Q1-2000:Q4 period
- evidence of **divergence** in the 2001:Q1-2010:Q4 period

Specifically, results show that

- neglecting spatial dependence might affect the results
- this is particularly evident in the 1981:Q1-1990:Q4 period