SPATIAL ECONOMIC ANALYSIS
https://doi.org/10.1080/17421772.2022.2095005

Routledge Taylor & Francis Group
RSA Regional Studies Association

OPEN ACCESS Check for updates

Distribution dynamics: a spatial perspective

Margherita Gerolimetto [a] and Stefano Magrini [a]

ABSTRACT
It is quite common in cross-sectional convergence analyses that data exhibit spatial dependence. Within the literature adopting the distribution dynamics approach, authors typically opt for spatial prefiltering. We follow an alternative route and propose a procedure based on an estimate of the mean function of a conditional density for which we develop a two-stage non-parametric estimator that allows for spatial dependence estimated via a spline estimator of the spatial correlation function. The finite sample performance of this estimator is assessed via Monte Carlo simulations. We apply the procedure that incorporates the proposed spatial non-parametric estimator to data on per capita personal income in US states and metropolitan statistical areas.

KEYWORDS
distribution dynamics, non-parametric estimation, spatial dependence

JEL C14, C21
HISTORY Received 10 May 2021; in revised form 25 May 2022

1. INTRODUCTION

Economic analyses are increasingly focusing on issues related to the spatial dimension of the problem under investigation. The importance of taking spatial dependence into account has clearly emerged since the seminal contributions of Paelinck and Klaassen (1979), Bartels and Ketellapper (1979) and Bennett (1979), which have stimulated a vast literature offering various tools to detect and treat spatial effects in empirical analyses.

The spatial dimension is certainly a relevant characteristic when studying convergence dynamics in per capita income across spatial units. In the traditional literature on convergence based on the regression approach there is now full awareness that neglecting spatial dependence may lead to biased and inefficient estimates. Drawing from the spatial econometrics literature, it is therefore common practice to resort to a spatial weight matrix, typically denoted by W, as a way to provide a parsimonious parametrization of interdependence relations between observations (LeSage & Pace, 2009). However, as recently emphasized by Stakhovych and Bijmolt (2009) and Corrado and Fingleton (2016), the choice of the spatial weights matrix is quite crucial as a misspecified W can bias the findings.

The issue has, however, received far less attention within the literature adopting the distribution dynamics approach. Within this approach, which analyses the evolution of the cross-sectional distribution of income by means of a conditional density function, also called stochastic kernel, the issue is tackled by adopting a spatial filtering technique involving an exogenously

CONTACT Stefano Magrini smagrini@unive.it
Department of Economics, Ca' Foscari University of Venice, Venice, Italy

assumed W before proceeding with the estimates. For example, Basile (2010) fits a spatial autoregressive model and employs residuals for subsequent analysis, while Fischer and Stumpner (2008) and Maza et al. (2010) employ a filtering approach based on the local spatial autocorrelation statistic Gi developed by Getis and Ord (1992). Clearly, however, pre-filtering is subject to the same critical remarks emphasized above. Particularly when knowledge of the spatial structure of the phenomenon under study is lacking and the risk of a misspecified W is high, we think this practice should be avoided: rather than filtering it away, information on spatial dependence could be extracted non-parametrically from the data and then included in the estimation process.

In this paper, we therefore propose a non-parametric technique that explicitly allows for spatial dependence in the distribution dynamics analysis, thus eliminating the need for pre-filtering while retaining statistical efficiency properties. As will be clarified in the following sections, the conditional density is commonly estimated using the kernel density estimator for which Hyndman et al. (1996) develop an adjustment procedure to deal with the so-called mean-bias issue. Here, we move from this idea by enriching the estimate of the conditional density with an estimate of the mean function that, in addition to Hyndman *et al.*'s original suggestion, allows for spatial dependence. To achieve this aim, we develop a two-step non-parametric regression estimator where the spatial dependence structure of the error terms is not a priori assumed; this piece of information is instead drawn from an estimate of the errors' spatial covariance matrix via a continuous non-parametric positive semi-definite consistent estimator of the spatial correlation function. In practical terms, the adopted non-parametric estimate of the spatial covariance frees the researcher from the need to assume that the structure of the interaction between spatial units is known.

From a methodological point of view, by proposing the spatial non-parametric (SNP) estimator we contribute to the existing literature in two aspects. First, we develop a general method for carrying out non-parametric regression that allows for spatial dependence within a fully non-parametric setting. Second, by incorporating SNP into Hyndman *et al.*'s mean bias adjustment procedure, we propose a tool for the analysis of distribution dynamics when data exhibit spatial dependence that is alternative to pre-filtering and eliminates the need for to be known to improve the efficiency of the estimates.

Finally, through this novel version of distribution dynamics approach, we contribute to the empirical literature by reconsidering the evidence on convergence dynamics across regional economies in the United States and shed some light on the consequences of neglecting spatial dependence. In particular, we analyse convergence between 1975 and 2008 using data on per capita personal income at two different spatial scales: a broader scale (48 conterminous states, excluding the District of Columbia) and a finer scale (380 metropolitan statistical areas – MSAs). In all, we find that both states and MSAs are characterized by a tendency towards convergence. However, the estimated extent and speed characterizing the convergence process depends on whether the presence of spatial dependence is allowed for.

The remainder of the paper is structured as follows. Section 2 recalls the distribution dynamics approach. Section 3 introduces the SNP estimator. Section 4 presents the application on per capita personal income data. Section 5 concludes.

2. DISTRIBUTION DYNAMICS

Distribution dynamics (Quah, 1993a, 1993b, 1996a, 1996b, 1997) is an approach to the analysis of convergence whose distinctive feature is to examine directly the evolution of the cross-sectional distribution of per capita income.[1]

In simple terms, consider a group of n economies and indicate with $Y_{j,t}$, defined on $\mathbb{R}$, per capita personal income of economy j at time (relative to the group average). Next, denote with $F_t(y)$ its distribution at time t and, assuming it admits a density, indicate this density with $f_t(y)$. Finally, assume that the dynamics of $F_t(y)$, or equivalently of, can be modelled as a first order process. As a result, the density prevailing at some future time $t+s$ is given by:

$$f_{t+s}(y') = \int_{-\infty}^{\infty} g_s(y'|y) f_t(y) dy, \tag{1}$$

where, $g_s(y'|y)$ is the s-period ahead density of y' conditional on y. Specifically, the conditional density function (1) maps the density at time into the density at time $t+s$ and therefore provides information both on the evolution of the external shape of the distribution and on intra-distributional dynamics between time t and time $t+s$.

A non-parametric (kernel) estimator of the conditional density in equation (1) can be obtained by dividing the estimator of the joint probability density function $f_{t,t+s}(y, y')$ by the estimator of the marginal probability density function $f_t(y)$:

$$\hat{g}_s(y'|y) = \frac{\hat{f}_{t,t+s}(y, y')}{\hat{f}_t(y)} \tag{2}$$

It is possible[2] to rewrite (2) as:

$$\hat{g}_s(y'|y) = \sum_{j=1}^{n} w_j(y) K_b(y' - Y_{j,t+s}) \tag{3}$$

where:

$$w_j(y) = \frac{K_a(y - Y_{j,t})}{\sum_{j=1}^{n} K_a(y - Y_{j,t})} \tag{4}$$

a and b are bandwidth parameters controlling the smoothness, $K_b(u) = b^{-1}K(u/b)$ is a scaled kernel function, $K(\cdot)$ is assumed to be a real value, integrable and non-negative even function[3]. Moreover, assuming that the conditional mean $E(y'|y) = M(y)$ exists, this can be estimated with the mean of the conditional density estimator in (3):

$$\hat{M}(y) = \int y' \hat{g}_s(y'|y) dy' = \sum_{j=1}^{n} w_j(y) Y_{j,t+s} \tag{5}$$

As further highlighted by Hyndman et al. (1996), the estimator in (5) is equivalent to the local constant regression estimator (LCE) of Nadaraya (1964) and Watson (1964) that is known to be affected by a large bias that is carried onto the corresponding estimator of the conditional density function. This bias, deriving from the estimated mean, is called mean-bias of a conditional density estimator.

As an alternative to reduce the mean-bias, Hyndman et al. (1996) propose a new class of conditional density estimators, defined as:

$$\hat{g}^*(y'|y) = \sum_{j=1}^{n} w_j(y) K_b(y' - Y^*_{j,t+s}(y)) \tag{6}$$

where $Y^*_{j,t+s}(y) = \hat{M}(y) + e_j - \sum_{i=1}^{n} w_i(y) e_i$ and, practically, $e_i = Y_{i,t+s} - \hat{M}(Y_{i,t})$, $i = 1, \ldots, n$. Expression (6) suggests that the mean-bias can be reduced by employing a non-parametric regression estimator with better bias properties than LCE. One such estimator is, for instance, the local linear estimator (LLE) (Cleveland, 1979; Fan & Gijbels, 1996):

$$\hat{M}(y) = \frac{\sum_{j=1}^{n} K_a(y - Y_{j,t})Y_{j,t+s}}{\sum_{j=1}^{n} K_a(y - Y_{j,t})} + (Y_{j,t} - \bar{Y}_w)\frac{\sum_{j=1}^{n} K_a(y - Y_{j,t})(Y_{j,t} - \bar{Y}_w)Y_{j,t+s}}{\sum_{j=1}^{n} K_a(y - Y_{j,t})(Y_{j,t} - \bar{X}_w)^2} \tag{7}$$

where:

$$\bar{Y}_w = \frac{\sum_{j=1}^{n} K_a(y - Y_{j,t})Y_{j,t}}{\sum_{j=1}^{n} K_a(y - Y_{j,t})}$$

This procedure is called mean-bias adjustment and, within a distribution dynamics setting, it effectively consists of an LLE estimate of the regression of per capita income at time $t + s$ on per capita income at time t. Since in empirical analyses of cross-sectional convergence units are unlikely to be independent, to improve the statistical properties of the estimator adopted in the mean-bias adjustment procedure and, to further increase the quality of the overall distribution analysis, spatial dependence should be handled. To tackle the issue without resorting to pre-filtering, we develop a spatially aware two-step procedure for non-parametric regression. We consider the very general case in which spatial dependence may arise from unmeasured variables that are related through space, aggregation of spatially correlated variables and systematic measurement error.[4]

3. NON-PARAMETRIC REGRESSION FOR SPATIALLY DEPENDENT DATA

3.1. Introduction

Non-parametric regression has now become quite a standard statistical tool when the functional form is possibly of an unknown type. Similarly to the parametric regression environment, non-parametric regression estimators generally assume i.i.d. error terms. In case of lack of independence, Robinson (1987, 2008, 2011) derives consistency and asymptotic distribution theory for the LCE in relation to various kinds of dependent data. Other authors (e.g., Lin & Carroll, 2000; Ruckstuhl et al., 2000; Wang, 2003; Xiao et al., 2003) study possible extensions of the non-parametric regression to a non-i.i.d. errors setting, where errors can be correlated and heteroskedastic. In all cases, however, a parametric structure for the dependence must be assumed beforehand and this might represent a serious limitation since, as highlighted by Martins-Filho and Yao (2009), most asymptotic results for the LCE and LLE in case of dependent errors are unfortunately contingent on the assumptions made on the covariance structure and it is not possible to generalize their application to different parametric structures. Stimulated by this lack of generality, Martins-Filho and Yao (2009) focus on estimators that, by incorporating the information contained in the error covariance structure, outperform, both asymptotically and in finite samples, traditional non-parametric ones. In particular, they propose a two-step procedure for the LLE in non-parametric regression under a general parametric error covariance and provide sufficient conditions for the asymptotic normality and efficiency relative to the traditional LLE. Along these lines, in what follows we describe a new two-step non-parametric regression estimator for spatially dependent data that has the advantage of not requiring a priori specification of the spatial covariance structure that, instead, is estimated non-parametrically.

3.2. Two-stage non-parametric regression

As emphasized at the end of Section 2, the procedure that is needed to improve the efficacy of the mean-bias adjustment by Hyndman et al. (1996) within a distribution dynamics analysis is a non-parametric (auto)regression for spatially dependent data. Note that the procedure we propose here is in fact more general than what strictly required for the mean-bias adjustment as it is exploitable on all occasions in which spatial dependence may represent a problem for a non-parametric regression analysis. Because of this generality, in this section we will adopt the common notation for univariate regressions where is the dependent variable and X is the independent one.

Let us consider the following non-parametric regression:

$$Y = M(X) + u \tag{8}$$

where the error term u is such that $E(u_i) = 0, \forall i = 1, \ldots, n$ and $E(u_i u_j) = \sigma^2 \rho_{ij}$ is the generic element of the autocovariance matrix V, so that ρ_{ij} is the ij-th element of the spatial autocorrelation matrix Ω. A commonly adopted approach (e.g., Anselin, 1988) to express the elements of Ω is through a direct, parsimonious representation of the dependence as some function of the distance separating sites, s_i and s_j. In such an instance, the spatial autocorrelation function is defined by:

$$\rho_{ij} = \rho(d_{ij}, \phi) \tag{9}$$

where d_{ij} is the distance between sites i and j, $\rho(\cdot)$ is a decaying function such that $\partial\rho/\partial d < 0$, $\phi \in \Phi$ as a $p \times 1$ vector of parameters in an open subset Φ of $\mathbb{R}^p$, $|\rho(d_{ij}, \phi)| \leq 1$. The hypothesis underpinning expression (9) is that data come from an isotropic second-order stationary process,[5] that is, the spatial autocorrelation function does not change through space, and it only depends on distance and not direction. This guarantees that the spatial autocorrelation matrix Ω is positive semi-definite and composed by elements (ρ_{ij}), obtained evaluating the function $\rho(\cdot)$ at observed distances across sites such that $\rho_{ii} = 1$.[6] Note that this framework is analogous to assumptions A3 and A6 in Martins-Filho and Yao (2009), who prove in theorems 2–4 the asymptotic normality of LLE in case of spatial dependence and the gain in efficiency characterizing a two-step LLE estimator that incorporates the information contained in the error covariance structure.

Given this set-up, to estimate $M(X)$ we propose the following procedure consisting of a sequence of steps:

(1) *Pilot fit*: a pilot estimate of $M(X)$ is obtained with an LLE estimator where the bandwidth, here denoted by h, is chosen following an optimal rule. The output of this step is the residuals set $\hat{u} = Y - \hat{M}(X)$.
(2) *Non-parametric covariance matrix estimation*: by means of a non-parametric estimate of the spatial autocorrelation function (presented in the next subsection) the spatial covariance matrix of u is consistently estimated through the residuals $\hat{u}$ coming from the first step. We denote this estimate by $\hat{V}$.
(3) *Final fit*: the procedure is fed with the information obtained from the estimate of the spatial covariance matrix $\hat{V}$ by running a modified regression where Y is replaced by a feasible quantity $\hat{Z}$ which is $\hat{Z} = \hat{M}(X) + \hat{L}^{-1}\hat{u}$, where L is obtained by taking the Cholevsky decomposition of $\hat{V}$ and $\hat{M}(X)$ and $\hat{u}$ derive from the pilot fit. At this stage, the error terms $\epsilon = \hat{L}^{-1}\hat{u}$ are spherical by construction and the final estimator is the LLE of the relationship of the new (feasible) regressand $\hat{Z}$ on X.

In other words, in its final step the procedure exploits the information contained in the error term correlation structure arising from the pilot fit, eventually yielding spherical errors. As for the bandwidths, note that undersmoothing in the pilot stage is required with respect to the modified regression. This is customary in the literature on two-stage non-parametric regression to avoid bias piling up (Martins-Filho & Yao, 2009).

3.3. Non-parametric estimation of the spatial covariance matrix

As said just above, in step 1 we need to get a consistent estimate of the residuals spatial covariance matrix; we derive $\hat{V}$ by means of a non-parametric estimate of the spatial correlation function by Bjørnstad and Falck (2001).

More in detail, Bjørnstad and Falck (2001) propose a continuous non-parametric positive semi-definite consistent estimator of the spatial correlation function. They build on the seminal work of Hall and Patil (1994) who develop the following kernel estimator of the spatial autocorrelation function for a stationary random field observed at points which are not assumed to be on a grid or lattice:[7]

$$\tilde{\rho}(d_{ij}) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} K_a(d_{ij})\hat{\rho}_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{n} K_a(d_{ij})} \tag{10}$$

where K_a is a kernel function with bandwidth a and $\hat{\rho}_{ij}$ is:

$$\hat{\rho}_{ij} = \frac{(z_i - \bar{z})(z_j - \bar{z})}{1/n \sum_{l=1}^{n} (z_l - \bar{z})^2} \tag{11}$$

where z is a generic variable, $\bar{z} = 1/n \sum_{l=1}^{n} z_l$ is its sample mean and d_{ij} are observed distances between sites i and j. Hall and Patil also demonstrate that the estimator in (10) can be tuned (by tuning a) so that $\tilde{\rho}(\cdot) \to \rho(\cdot)$ as $n \to \infty$ for any smooth functional form of $\rho(\cdot)$, that is with continuous first and second derivatives.

Starting from the estimator in (10), Bjørnstad and Falck (2001) express the kernel function in the form of a cubic B-spline.[8] The advantage in using the B-spline is in that this smoother adapts better to irregularly spaced data and produces a consistent estimate of the correlation function (Hyndman & Wand, 1997). Moreover, it has been shown that fixing the degree of smoothing using cross validation (see Green & Silverman, 1994; and Hastie et al., 2009, for more details) guarantees results with asymptotic properties.

Finally, since the estimator of $\rho(\cdot)$ must be not only pointwise consistent but also positive semi-definite, which is not necessarily guaranteed by the estimator $\tilde{\rho}(\cdot)$ in equation (10), Bjørnstad and Falck (2001) resort to a Fourier filter method (Hall et al., 1994). Based on Bochner's theorem, this method works as follows: firstly the Fourier transform of $\tilde{\rho}(\cdot)$ is calculated, then all negative excursions of the transformed function are set to zero and, last, a non-parametric positive semi-definite estimate of the spatial correlation function is obtained by back-transformation.

3.4. Monte Carlo study

We conduct an extensive Monte Carlo experiment[9] to show the finite sample performance of our procedure, denoted by SNP, in comparison with a traditional LLE regression that ignores the presence of spatial dependence, denoted by NP. The purpose therefore is to investigate the effective improvement in regression estimation results when spatial dependence is taken into account.

The Monte Carlo experiment is carried out considering the following non-linear specifications:

(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) $M(x) = \left(0.3\sqrt{2\pi}\right)^{-1} exp(-2(X - 0.5)^2)$

for the model:

$$Y = M(X) + u$$
$$u = \lambda Wu + \epsilon$$

The shapes of the functional forms are depicted in Figure 1.

The simulated data set length is $N = 50$, 100, 200 and the number of Monte Carlo replications per experiment is 1000. The regressor is drawn from a uniform distribution, $X \sim U(0, 1)$, while ε is generated as a vector of normally distributed random variables, $\epsilon \sim N(0, \sigma^2)$, where σ^2 is set to obtain three levels for the pseudo-R^2 (0.3, 0.5, 0.7). Starting from a Euclidean distance matrix obtained from randomly generated coordinates, two W matrices have been considered: 10%nearest-neighbours ($W1$) and contiguity from Voronoi tessellation ($W2$). Finally, λ takes on three values (0.0, 0.4, 0.8), giving us a total of 180 ($3 \times 5 \times 3 \times 2 \times 2$) experiments.

As said above, we employ two estimation methods: the traditional non-parametric estimator (NP) and our procedure (SNP), both in the form of an LLE. For all simulations, we use the gaussian kernel. As for the bandwidths, we adopt cross-validation and direct plug-in.[10] In the estimate of the spatial autocorrelation function, the smoothing parameter is selected by cross-validation.[11]

As is common in this type of simulation experiments, estimators' performance is measured by calculating the median across replications of the mean integrated squared error (MISE) obtained in each replication. Tables 1 and 2 show the median MISE of SNP for both W matrices.

A direct comparison of the relative performance of the two estimators is then carried out through the ratio between the median MISE of SNP with respect to the median MISE of NP. These results are reported in Tables 3 and 4.[12]

Overall, the performance of SNP is quite good as median ratios are below 1 in nearly all cases, with no appreciable differences across the considered functional forms and spatial weights matrices employed in the data-generating process. Median ratios are closer to 1 when spatial dependence is absent ($\lambda = 0.0$) and for the smallest sample size ($N = 50$), while they display significant reductions as the strength of spatial dependence and the size of the sample increase. In particular, the SNP procedure visibly outperforms the traditional LLE when reaches 0.8 and $N = 200$, obtaining median values of the MISE that are approximately 25% smaller in several cases. Finally, median ratios appear to be somewhat lower when the direct plug-in bandwidth is employed.

Some attention should be given to the case when spatial dependence is absent ($\lambda = 0.0$). Interestingly, also in such case median ratios are smaller than 1, thus revealing also in this case the capability of SNP of outperforming NP. To show this more clearly, and understand the reasons behind this outcome, we present some more in-depth results through a further simulation experiment that concentrates on the case $\lambda = 0$ (over all non-linear functional forms).

We start by excluding that this behaviour depends on the double smoothing implicit in the two-step nature of SNP. Simulations results (presented in Appendix A2 in the supplemental data online) confirm what the theory suggests: the use of SNP without performing its intermediate step (i.e., the spline estimate of the correlation function) which amounts to repeating just two subsequent NP smooths (denoted by NP2s) produces a worse performance compared with the

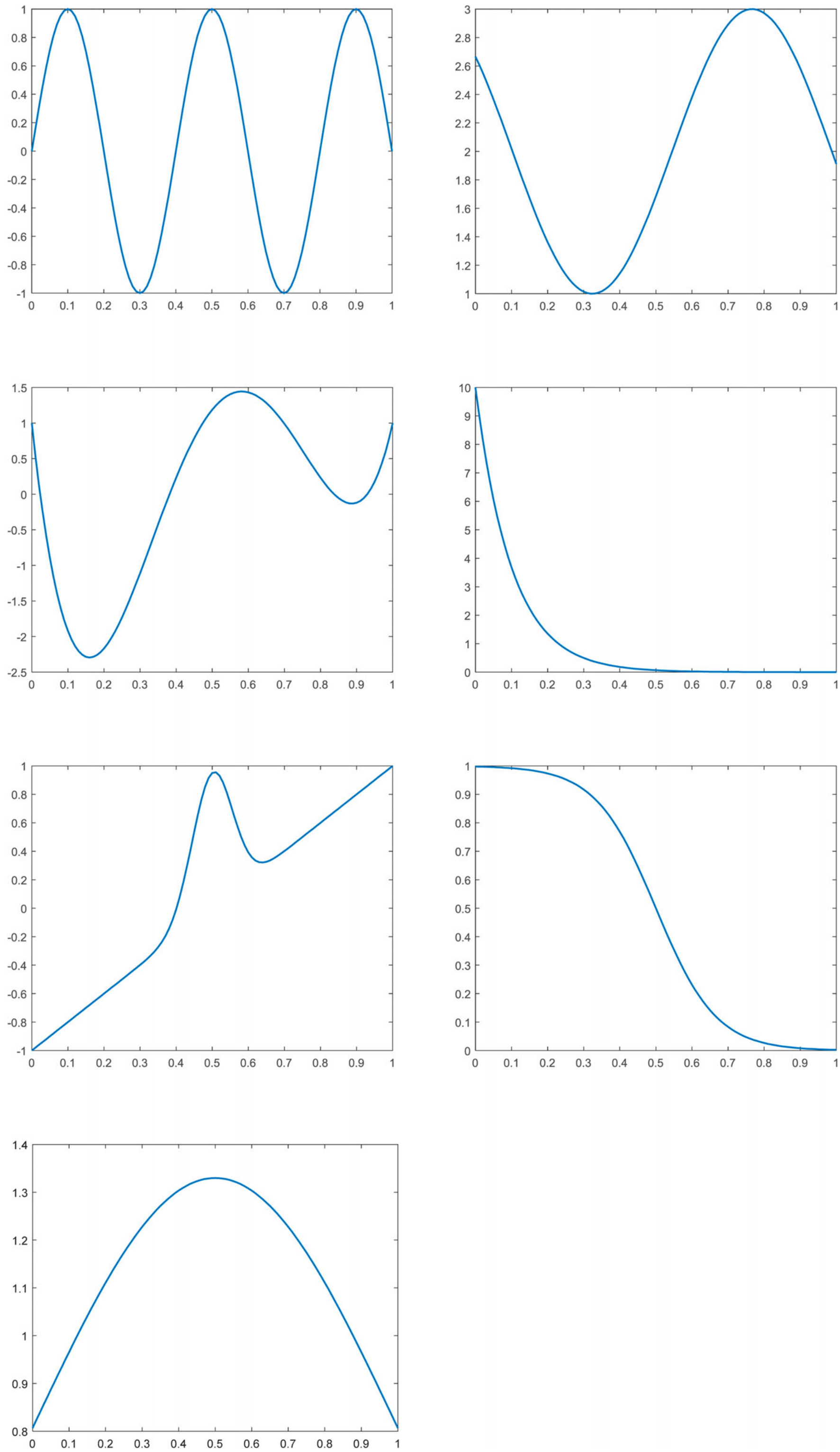


Figure 1. Monte Carlo experiment: functional forms.

Table 1. Monte Carlo: median SNP MISE – cross-validation minimization – $W1$.

		A			B			C		
Pseudo-R^2	n	**0**	**0.4**	**0.8**	**0**	**0.4**	**0.8**	**0**	**0.4**	**0.8**
0.7	50	481.05	478.75	479.43	249.70	243.20	235.02	740.19	717.86	693.85
0.7	100	461.74	459.69	458.06	262.76	252.35	236.30	726.38	703.29	689.49
0.7	200	445.67	441.02	444.65	267.33	251.98	222.84	691.62	646.89	639.28
0.5	50	486.93	485.34	484.75	257.99	250.39	247.86	757.00	737.61	736.73
0.5	100	466.87	460.75	463.86	262.22	244.80	243.85	737.17	714.29	718.03
0.5	200	444.82	443.47	449.32	265.93	239.61	234.41	681.74	649.21	645.81
0.3	50	493.85	494.58	493.27	271.30	271.61	267.38	803.12	789.75	794.72
0.3	100	475.25	470.00	471.01	259.14	252.10	257.58	745.25	735.34	735.24
0.3	200	452.10	451.00	455.58	270.33	238.22	240.88	693.96	661.51	663.36

		D			E			F		
		0	**0.4**	**0.8**	**0**	**0.4**	**0.8**	**0**	**0.4**	**0.8**
0.7	50	1029.96	991.36	1069.05	68.23	66.72	66.45	11.18	11.16	11.07
0.7	100	1109.95	1057.24	1145.50	63.30	62.57	63.78	10.90	10.52	10.32
0.7	200	1156.58	1039.20	1132.69	60.07	58.83	59.92	10.66	10.03	9.57
0.5	50	1119.25	1066.50	1146.67	73.69	72.86	72.29	12.69	12.58	12.66
0.5	100	1153.28	1071.10	1185.15	65.75	65.23	66.04	11.53	11.47	11.39
0.5	200	1173.81	1083.03	1196.45	60.89	60.29	61.99	11.07	10.86	10.62
0.3	50	1290.36	1296.14	1372.36	82.07	82.38	81.48	16.71	16.71	16.43
0.3	100	1250.65	1198.36	1275.75	71.75	71.24	70.90	13.10	12.94	12.73
0.3	200	1262.09	1138.12	1227.80	65.58	64.18	65.37	11.75	11.77	11.79

		G		
		0	**0.4**	**0.8**
0.7	50	13.36	12.42	13.09
0.7	100	12.53	11.80	12.57
0.7	200	12.01	9.46	10.88
0.5	50	13.27	12.02	13.58
0.5	100	11.85	11.47	13.01
0.5	200	12.83	9.55	11.60
0.3	50	14.59	13.98	14.94
0.3	100	12.61	11.57	12.97
0.3	200	12.94	10.23	12.24

Table 2. Monte Carlo: median SNP MISE – cross-validation minimization – *W*2.

		A			**B**			**C**		
Pseudo-R^2	***n***	**0**	**0.4**	**0.8**	**0**	**0.4**	**0.8**	**0**	**0.4**	**0.8**
0.7	50	481.05	477.32	478.18	249.70	246.03	237.17	740.19	722.60	696.62
0.7	100	461.74	458.54	455.92	262.76	250.48	230.27	726.38	701.02	702.30
0.7	200	445.67	441.47	445.50	267.33	220.39	219.44	691.62	633.16	641.16
0.5	50	486.93	484.85	484.40	257.99	249.31	246.12	757.00	739.24	725.66
0.5	100	466.87	463.37	460.59	262.22	240.75	237.67	737.17	721.07	725.55
0.5	200	444.82	443.55	448.17	265.93	218.05	225.89	681.74	634.36	646.97
0.3	50	493.85	491.32	494.43	271.30	268.71	260.02	803.12	792.51	783.50
0.3	100	475.25	468.84	469.16	259.14	246.64	255.08	745.25	724.71	740.52
0.3	200	452.10	448.21	454.07	270.33	219.84	235.81	693.96	651.13	664.03

		D			**E**			**F**		
		0	**0.4**	**0.8**	**0**	**0.4**	**0.8**	**0**	**0.4**	**0.8**
0.7	50	1029.96	987.07	1029.05	68.23	66.84	67.63	11.18	11.11	10.93
0.7	100	1109.95	1030.67	1125.36	63.30	61.98	63.11	10.90	10.39	10.23
0.7	200	1156.58	1007.75	1126.83	60.07	58.21	59.89	10.66	9.38	9.43
0.5	50	1119.25	1063.28	1100.08	73.69	72.76	72.02	12.69	12.65	12.18
0.5	100	1153.28	1056.76	1168.37	65.75	63.91	65.77	11.53	11.47	11.30
0.5	200	1173.81	1041.34	1183.68	60.89	59.49	61.34	11.07	10.75	10.56
0.3	50	1290.36	1264.54	1296.48	82.07	82.03	80.58	16.71	16.25	15.58
0.3	100	1250.65	1205.57	1259.14	71.75	71.34	70.51	13.10	13.04	12.44
0.3	200	1262.09	1097.91	1213.41	65.58	64.15	64.96	11.75	11.83	11.73

		G		
		0	**0.4**	**0.8**
0.7	50	13.36	12.43	12.74
0.7	100	12.53	11.49	12.36
0.7	200	12.01	8.71	10.74
0.5	50	13.27	12.29	13.00
0.5	100	11.85	11.23	13.13
0.5	200	12.83	8.82	11.39
0.3	50	14.59	14.04	14.15
0.3	100	12.61	11.71	13.36
0.3	200	12.94	9.20	11.91

Table 3. Monte Carlo: median MISE ratios – cross-validation bandwidth – $W1$.

		A			B			C		
Pseudo-R^2	***n***	**0**	**0.4**	**0.8**	**0**	**0.4**	**0.8**	**0**	**0.4**	**0.8**
0.7	50	0.98	0.98	0.98	0.93	0.91	0.88	0.93	0.90	0.88
0.7	100	0.97	0.96	0.96	0.94	0.90	0.84	0.90	0.88	0.87
0.7	200	0.96	0.95	0.95	0.94	0.88	0.78	0.88	0.83	0.82
0.5	50	0.98	0.98	0.98	0.94	0.91	0.90	0.94	0.91	0.91
0.5	100	0.97	0.96	0.97	0.93	0.87	0.86	0.92	0.89	0.89
0.5	200	0.95	0.95	0.96	0.92	0.84	0.82	0.87	0.82	0.82
0.3	50	0.99	0.98	0.98	0.96	0.95	0.93	0.97	0.95	0.94
0.3	100	0.98	0.97	0.97	0.91	0.88	0.89	0.91	0.90	0.90
0.3	200	0.96	0.96	0.96	0.92	0.82	0.83	0.87	0.83	0.83
		D			E			F		
		0	**0.4**	**0.8**	**0**	**0.4**	**0.8**	**0**	**0.4**	**0.8**
0.7	50	0.89	0.85	0.90	0.96	0.94	0.94	0.98	0.98	0.96
0.7	100	0.81	0.78	0.86	0.94	0.93	0.95	0.97	0.94	0.92
0.7	200	0.80	0.72	0.78	0.92	0.89	0.92	0.94	0.89	0.85
0.5	50	0.92	0.88	0.95	0.99	0.99	0.97	1.00	0.98	0.96
0.5	100	0.84	0.79	0.88	0.95	0.94	0.95	0.98	0.98	0.98
0.5	200	0.79	0.72	0.81	0.91	0.90	0.93	0.97	0.95	0.93
0.3	50	0.95	0.96	0.97	1.00	1.01	0.96	1.01	1.00	0.91
0.3	100	0.88	0.84	0.91	0.99	0.98	0.98	0.99	1.00	1.00
0.3	200	0.83	0.75	0.83	0.95	0.93	0.95	0.99	0.99	0.99
		G								
		0	**0.4**	**0.8**						
0.7	50	0.82	0.77	0.81						
0.7	100	0.76	0.72	0.77						
0.7	200	0.76	0.60	0.69						
0.5	50	0.82	0.76	0.82						
0.5	100	0.72	0.70	0.79						
0.5	200	0.80	0.59	0.73						
0.3	50	0.86	0.82	0.87						
0.3	100	0.76	0.70	0.82						
0.3	200	0.79	0.62	0.75						

simple NP. We thus infer that the superior performance offered by SNP even when $\lambda = 0$ should be attributed to the intermediate step in which the covariance matrix is estimated.

To confirm this, on the grounds that, in theory, when residuals are spherical the correlation matrix corresponds to the identity matrix, we calculate a measure of distance between the identity matrix and each of the following 4 correlation matrices:

Table 4. Monte Carlo: median SNP MISE – cross-validation minimization – *W*2.

		A			B			C		
Pseudo-R^2	***n***	**0**	**0.4**	**0.8**	**0**	**0.4**	**0.8**	**0**	**0.4**	**0.8**
0.7	50	0.98	0.97	0.97	0.93	0.92	0.88	0.93	0.91	0.88
0.7	100	0.97	0.96	0.96	0.94	0.89	0.82	0.90	0.87	0.87
0.7	200	0.96	0.95	0.95	0.94	0.78	0.77	0.88	0.81	0.81
0.5	50	0.98	0.98	0.98	0.94	0.91	0.90	0.94	0.91	0.90
0.5	100	0.97	0.96	0.96	0.93	0.85	0.84	0.92	0.89	0.90
0.5	200	0.95	0.95	0.95	0.92	0.76	0.79	0.87	0.81	0.82
0.3	50	0.99	0.98	0.98	0.96	0.95	0.92	0.97	0.95	0.94
0.3	100	0.98	0.97	0.97	0.91	0.86	0.89	0.91	0.88	0.90
0.3	200	0.96	0.95	0.96	0.92	0.76	0.81	0.87	0.81	0.83
		D			**E**			**F**		
		0	**0.4**	**0.8**	**0**	**0.4**	**0.8**	**0**	**0.4**	**0.8**
0.7	50	0.89	0.85	0.87	0.96	0.94	0.95	0.98	0.98	0.96
0.7	100	0.81	0.77	0.83	0.94	0.92	0.94	0.97	0.92	0.91
0.7	200	0.80	0.70	0.78	0.92	0.89	0.92	0.94	0.83	0.83
0.5	50	0.92	0.88	0.91	0.99	0.98	0.97	1.00	0.99	0.97
0.5	100	0.84	0.77	0.85	0.95	0.93	0.95	0.98	0.99	0.97
0.5	200	0.79	0.70	0.80	0.91	0.89	0.93	0.97	0.93	0.92
0.3	50	0.95	0.93	0.97	1.00	1.01	0.99	1.01	1.01	0.95
0.3	100	0.88	0.85	0.89	0.99	0.98	0.97	0.99	1.01	0.97
0.3	200	0.83	0.73	0.83	0.95	0.93	0.94	0.99	0.98	0.95
		G								
		0	**0.4**	**0.8**						
0.7	50	0.82	0.78	0.79						
0.7	100	0.76	0.69	0.75						
0.7	200	0.76	0.55	0.68						
0.5	50	0.82	0.76	0.80						
0.5	100	0.72	0.68	0.79						
0.5	200	0.80	0.55	0.72						
0.3	50	0.86	0.82	0.83						
0.3	100	0.76	0.70	0.80						
0.3	200	0.79	0.56	0.73						

- The correlation matrix of the simulated ϵ for the Monte Carlo experiment when $\lambda = 0$.
- The correlation matrix of the residuals of the traditional NP regression.
- The correlation matrix of the residuals of the pilot fit of the SNP procedure (i.e., a traditional NP regression with an undersmoothed bandwidth).
- The correlation matrix of the residuals resulting from the SNP procedure.

The first distance, denoted by d_ε, is adopted as a benchmark to evaluate the others (d_{NP}, d_{SNPp} and d_{SNP}, respectively). The adopted measure of distance is the Frobenius norm[13] and for each of the four distances, the median across 1000 replications is proposed. Results, reported in Appendix A3 in the supplemental data online, show that, in spite of all distances being rather small, hence implying that all correlation matrices are close to the identity, yet the SNP residuals correlation matrix is definitely the closest one. Despite not being statistically significant (as confirmed by Moran'I test results available upon request), the structure generated in the computation negatively affects the performance (in terms of MISE) of the NP estimator; on the other hand, the intermediate spline estimate of the correlation function has the effect of cleaning the residuals from spurious patterns and determines the gain in performance. This gain is particularly more evident the more complex is the functional form of the data generating process.

Summing up, SNP provides a general improvement in terms of performance with respect to its a-spatial counterpart. When there is lack of information about the spatial structure of the phenomenon under study, SNP can be a valid option also in comparison to a parametric regression whose results are known to be sensitive to the choice of the spatial weights matrix. One of the instances in which this lack of information is a commonplace, is in the study of convergence across a cross-section of economies through the distribution dynamics approach, and this is what motivates this paper.

5. EMPIRICAL ANALYSIS

We study economic convergence across the US economy employing data on (the logarithm of) per capita personal income at two different spatial scales: a broader scale (48 conterminous states, excluding the District of Columbia) and a finer scale (380 MSAs).[14]

As shown in Magrini et al. (2015) and Gerolimetto and Magrini (2017), since regional disparities in the United States follow a distinct cyclical pattern in the short run, results from convergence analysis could be affected by sizeable distortions when the period under scrutiny includes incomplete cycles. In order to avoid them, we adopt a composite strategy. First, we follow the approach described by Gerolimetto and Magrini (2017) and begin the analysis by extracting the trend from each of the per capita personal income series using the Hodrick–Prescott filter (Hodrick & Prescott, 1997). As suggested by Ravn and Uhlig (2002), the value for the parameter that controls the degree of smoothness of the estimated trend is set to 6.25. Once estimated the trends, in line with Magrini et al. (2015), we select two points in time that correspond to similar phases of the business cycle: the trough that occurred during the First Oil Crisis and the trough that occurred during the Great Recession[15]. As a result, the years select are 1975 and 2008 and we then study convergence by applying the distribution dynamics approach to corresponding data on the extracted trends, that is, in terms of the notation adopted in Section 2, t is 1975 and $t+s$ is 2008. Operatively, we estimate the stochastic kernel and then calculate the corresponding ergodic distribution, that is, the limiting distribution whose external shape does not change over time while allowing for intra-distribution movements according to the stochastic kernel. In particular, following Johnson (2005), we calculate the ergodic distribution, denoted by $f_\infty(y')$, corresponding to a given stochastic kernel by solving:

$$f_\infty(y') = \int_{-\infty}^{\infty} g_s(y'|y) f_\infty(y)\,dy \qquad (12)$$

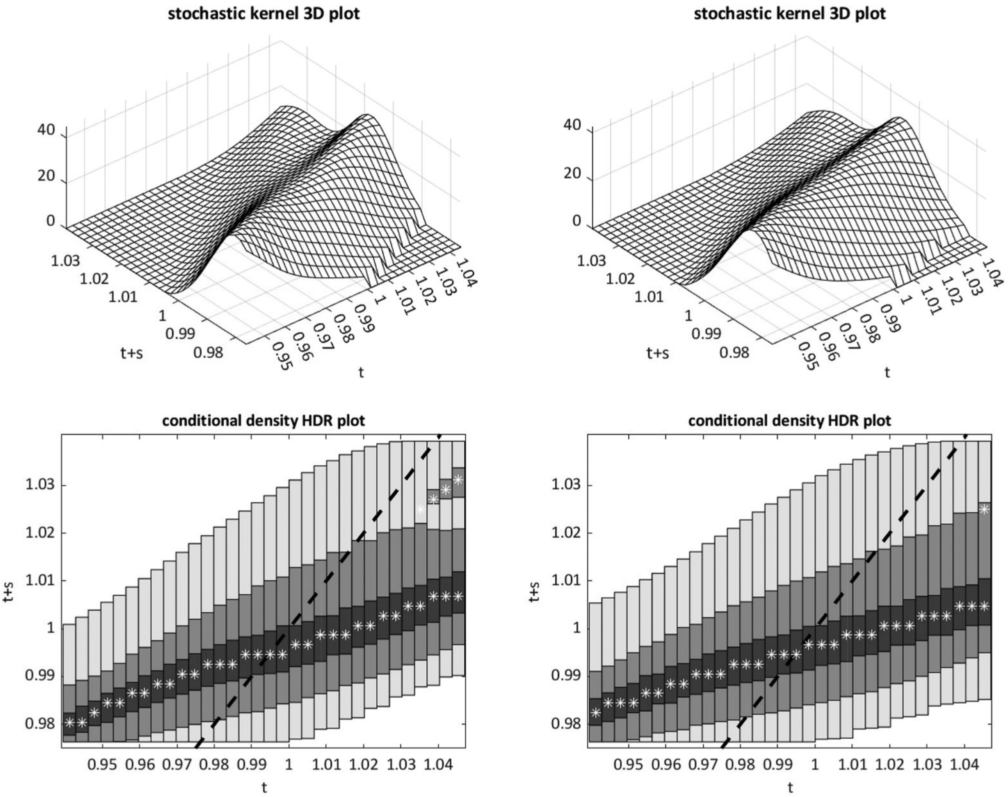
To emphasize the way in which spatial dependence can affect the estimates, we visually compare stochastic kernels obtained using both NP and SNP estimators in the mean's function adjustment procedure.[16] In these plots, a clockwise (counter-clockwise) rotation of the estimated probability mass from the diagonal[17] indicates that a process of convergence (divergence) has occurred during the analysed period. To get an idea of the speed with which distributions evolve and reach a stationary shape, we resort to the concept of asymptotic half-life of the chain (Shorrocks, 1978), that is the amount of time taken to cover half the distance to the ergodic distribution. In addition, we report Moran's I index of spatial dependence (and corresponding p-value based on the randomization assumption and using a K-nearest neighbour row-standardized matrix, with K corresponding to the 10% of the total number of observations) on filtered data as well as on NP or SNP residuals of the mean function estimates; when the test suggests the presence of spatial dependence in the residuals, a Moran significance maps is also displayed to show locations with a significant local Moran statistic (at the 10% significance level). Finally, we report two dispersion measures, the coefficient of variation and interquartile range, for 1975, 2008 and ergodic distributions.

We begin the analysis from the states. The upper panels in Figure 2 show the three-dimensional plots of the estimated stochastic kernels – via the traditional NP on the left and SNP on the right – while the lower panels show the corresponding high-density region (HDR) plots in which the vertical strips represent conditional densities for a specific value in the initial year dimension and, for each strip, darker to lighter areas display the 25%, 70% and 95%HDRs. Using either estimator, the estimated probability mass displays an evident clockwise rotation from the diagonal thus indicating that a robust process of convergence in per capita personal income levels across US states occurred during the analysed period. There are, however, a couple of minor differences in the estimates that should be underlined. First, in the estimate obtained through the traditional NP estimator, the evident probability mass located in the top-right corner of the HDR plot indicates that the economies characterized by a high initial level of per capita personal income (relative to the sample average) show a higher probability to maintain a sizeable positive gap over the other economies. Second, in the case of SNP the darkest section of the probability mass (identifying the top 25% of each conditional probability) as well as the sequence of modes (the asterisks) appear to lay on a flatter line than in the case of NP. These features suggest that extent and speed of the convergence process is lower according to the NP estimator.

Table 5 confirms the visual impressions. The variation coefficient and interquartile range values corresponding to the ergodic distributions are substantially lower than those corresponding to the 1975 distribution, and in line with the values corresponding to the 2008 distribution, suggesting that the convergence process has come to its end towards the final part of the analysed period. This is confirmed by the estimate of the speed with which the distribution approaches its stationary shape: the amount of time needed in order to cover half of the distance to the ergodic distribution is between 64% (NP) and 54% (SNP) of the period length. This 15% difference in the estimated half-life values and, to a lesser extent, the differences in dispersion statistics affirms that, when the traditional NP estimator is employed, estimated speed and extent of convergence is lower.

As discussed in the previous sections, the differences in the results could arise because spatial dependence is not properly allowed for by NP. To shed some light on this aspect, we first quantify the extent to which spatial dependence affects the data and the residuals of the regression for the mean function estimation. Table 6 reports the Moran's I index values. Contingent on this choice of W, we find strong evidence of spatial dependence in states' per capita personal income at the end of the period (2008) as well as in the residuals of the NP regression; in contrast, we find no evidence of spatial dependence in per capita personal income in 1975 nor in the residuals of the SNP regression.

Concentrating on the NP regression residuals, the significance map in Figure 3 shows the locations with significant local Moran statistics: a group of states on the north-east corner of



(a) NP (b) SNP

Figure 2. Distribution dynamics (states). Note: Estimates use a nearest-neighbour bandwidth in the initial year (1975) dimension (span = 0.7), a normal scale (Silverman, 1986) bandwidth in the final year (2008) dimension and a Gaussian kernel. The (fixed) bandwidth for both NP and SNP estimates is chosen using cross-validation (Hurvich et al., 1998). In the HDR plot, the dashed line represents the main diagonal, the asterisk the modes.

the country share significantly higher residual values; on the western side, Arizona is characterized by low values and is surrounded by states with high values; finally, Wyoming has high values but is surrounded by states with low values.

Table 5. Distribution dynamics statistics (states).

Series	Variation coefficient	Interquartile range
1975	0.0245	0.0361
2008	0.0135	0.0173
Ergodic – NP	0.0136	0.0182
Ergodic – SNP	0.0134	0.0180

	Half-life iterations
Ergodic – NP	0.6374
Ergodic – SNP	0.5406
Δ (%)	15.19%

In addition, Figure 4 reports the spatial correlogram estimated within the SNP procedure in the left panel and, in the right one, compares the shapes of the mean functions estimated by SNP (continuous line) and NP (dashed line). The spatial correlogram exhibits the typical decaying shape indicated by the theoretical literature in the presence of a correlation pattern. This feature then translates into different mean function estimates: specifically, compared with the mean function estimated through NP, the one obtained through SNP is characterized by a clockwise rotation.

To sum up, the distribution dynamics analysis suggests that the 1975–2008 period is characterized by a general reduction in cross-sectional per capita personal income disparities across US states. At the same time, the analysis also shows that, if not allowed for, spatial dependence may affect the estimate of the mean function; this, in the analysed case, leads to an underestimation of the speed and extent of the convergence process.

We now turn to the finer set of spatial units: MSAs. On the whole, also at this spatial scale the estimated conditional probabilities, depicted in Figure 5, display an evident clockwise rotation with respect to the 45° line, indicative of a process of economic convergence. There are however some more subtle features that must be emphasized. First, using both estimators, the rotation of the probability mass is less accentuated than in the states' case (Figure 2), suggesting that the speed of convergence is slower for MSAs. Second, the shape of the probability mass estimated using SNP appears remarkably non-linear and follows more closely the 45° along both tails; this, in turn, suggests that convergence speed is lower when the mean function is estimated through the SNP procedure.

These results are confirmed by the statistics reported in Table 7. The dispersion measures calculated for the ergodic distributions are essentially in line and, on the other hand, their values are substantially lower to those calculated in 1975. As for the speed with the distributions approach their steady-state shapes, this is approximately 20% lower when estimated through SNP as the amount of time needed to reach half-way to the ergodic distribution is 100% of the length of the analysed period (i.e., 33 years) compared with 80% estimated using the traditional NP.

Spatial dependence statistics reported in Table 8 show strong evidence of spatial dependence in the data as well as in the residuals of the NP regression; in contrast, there is no evidence of spatial dependence in the residuals of the SNP regression.

Given the presence of spatial dependence in NP regression residuals, Figure 6 shows the locations with significant local Moran statistics. In the case of MSAs, the maps indicate a more complex phenomenon from a spatial point of view, with several hotspots scattered all over the US territory.

The extent and structure of spatial dependence can be studied further through the spatial correlogram estimated within the SNP procedure and drawn in the left panel of Figure 7. Compared with the states' case, and coherently with the phenomena depicted in the Moran significance maps, the correlogram from the MSAs decays at a sensibly faster rate. The consequence of this feature on the estimated mean functions is portrayed in the right panel of Figure 7. The shape of the mean function estimated through SNP is clearly non-linear and follows more closely the 45° along both tails. As suggested above in the discussion of the conditional probability

Table 6. Spatial dependence statistics (states).

Series	Moran's *I*	*p*-value
1975	0.1029	0.1258
2008	0.3881	0.0000
Residuals – NP	0.3905	0.0000
Residuals – SNP	−0.0942	0.3695

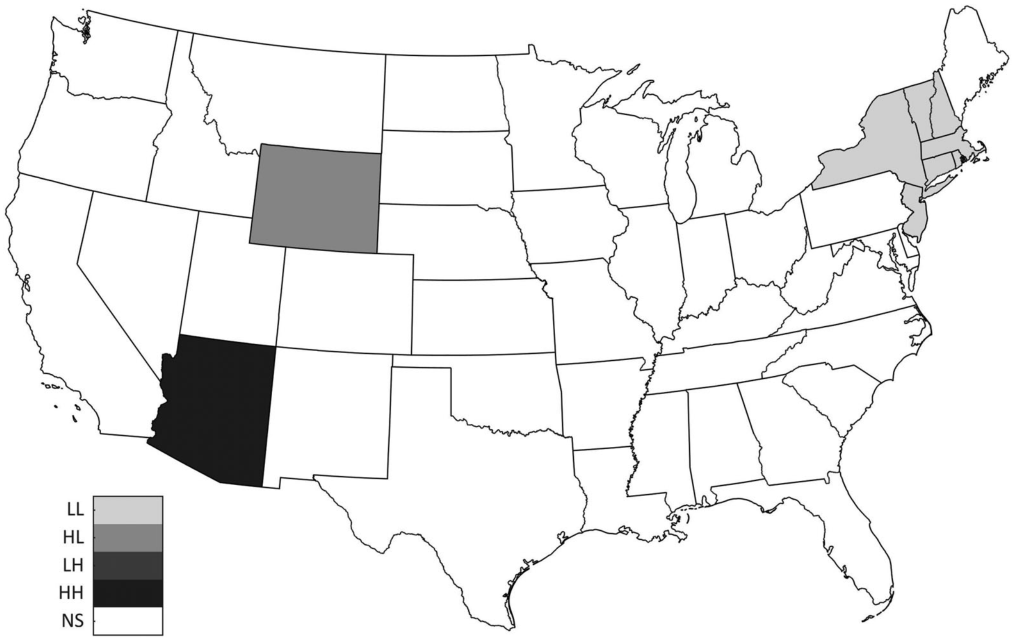
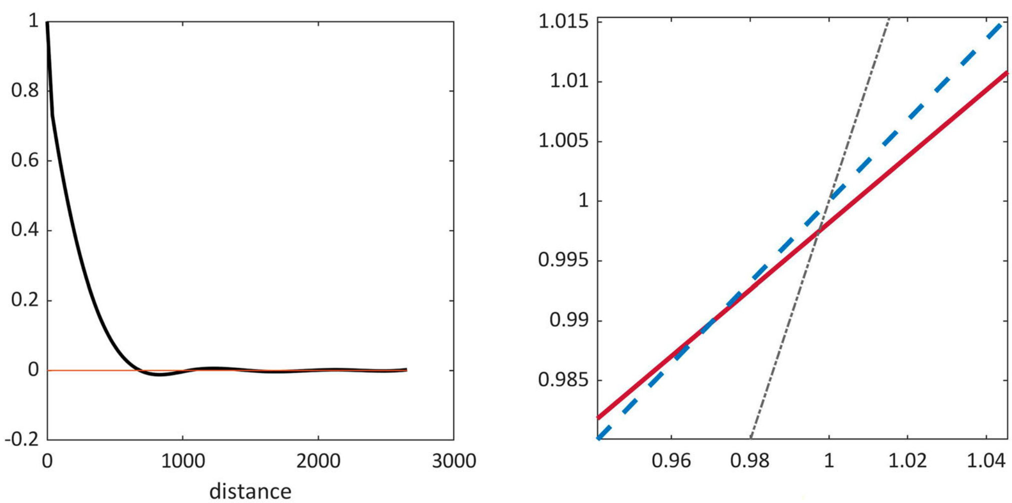
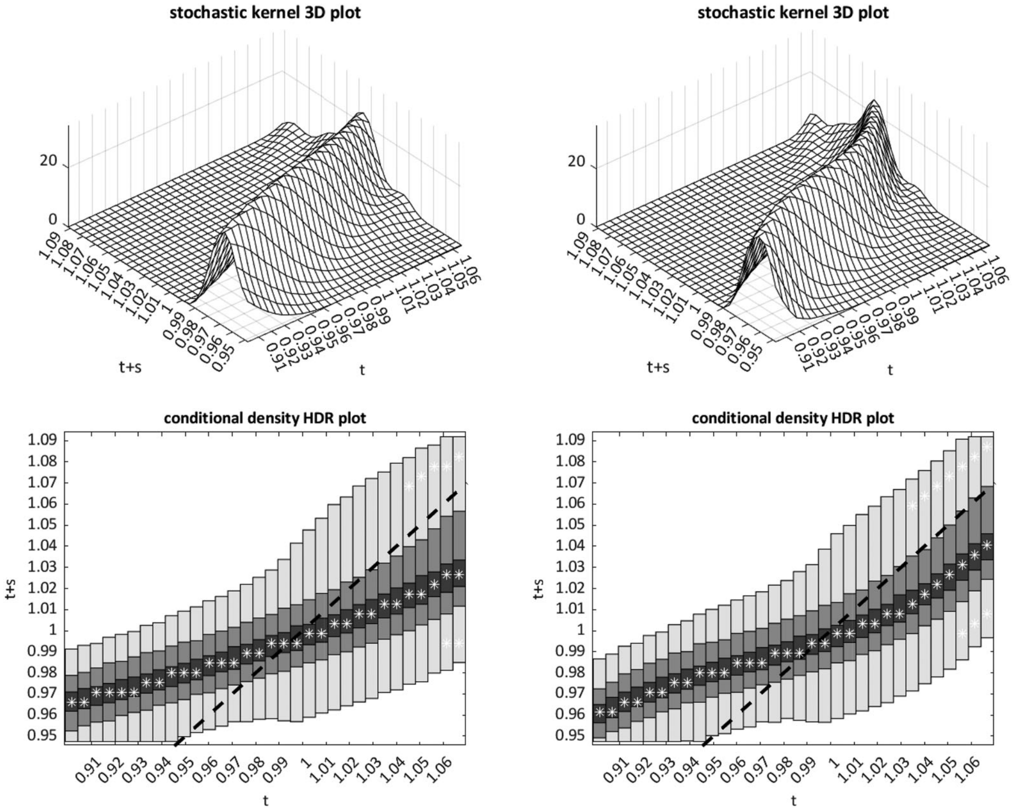


Figure 3. Moran significance map (states). Note: The map shows statistically significant ($p < 0.10$) local Moran's I-statistics using a 10% nearest-neighbours, row-standardized W matrix.



(a) spatial correlogram (SNP) (b) mean function estimates

Figure 4. Spatial correlogram and mean function estimates (states). Note: In the left panel, the number of knots in the estimate of the spline correlogram is chosen using cross-validation. In the right panel, the continuous line represents the SNP estimate, the dashed line represents the NP estimate, and the dash–dotted line represents the 45° line. The (fixed) bandwidth for both estimates is chosen using cross-validation (Hurvich et al., 1998).



(a) NP (b) SNP

Figure 5. Distribution dynamics (MSAs). Note: Estimates use a nearest-neighbour bandwidth in the initial year (1975) dimension (span = 0.7), a normal scale (Silverman, 1986) bandwidth in the final year (2008) dimension and a Gaussian kernel. The (fixed) bandwidth for both NP and SNP estimates is chosen using cross-validation (Hurvich et al., 1998). In the HDR plot, the dashed line represents the main diagonal, the asterisk the modes.

Table 7. Distribution dynamics statistics (MSAs).

Series	Variation coefficient	Interquartile range
1975	0.0238	0.0299
2008	0.0165	0.0193
Ergodic – NP	0.0158	0.0196
Ergodic – SNP	0.0159	0.0197
	Half-life iterations	
Ergodic – NP	0.8451	
Ergodic – SNP	1.0220	
Δ (%)	–20.93%	

Table 8. Spatial dependence statistics (MSAs).

Series	Moran's *I*	*p*-value
1975	0.1811	0.0000
2008	0.0913	0.0000
Residuals – NP	0.1282	0.0000
Residuals – SNP	0.0074	0.3400

estimates, the different shape of the estimated mean functions confirms that, in the case of MSAs, convergence speed is lower when the estimated structure of spatial dependence is allowed for through the use of the SNP procedure.

6. CONCLUSIONS

In this paper, we have proposed a spatially aware distribution dynamics approach to the analysis of economic convergence. The special feature of the approach is that the non-parametric estimate of the mean function underlying the stochastic kernel is enriched by an additional step to exploit the information about the spatial structure among units that might affect estimates. This structure is not a priori assumed in the form of a given spatial weight matrix, as it is common in the spatial econometric literature, but it is instead drawn by a non-parametric estimate of the spatial correlation function. The outcome is the development of a two-step non-parametric estimator that allows for spatial dependence aimed at avoiding the bias, recently emphasized in the literature, determined by a misspecified W matrix.

The two-step SNP estimator is by all means a valuable tool in itself. We conducted an extensive Monte Carlo experiment showing that SNP outperforms the traditional, a-spatial non-parametric estimator in terms of mean integrated squared error.

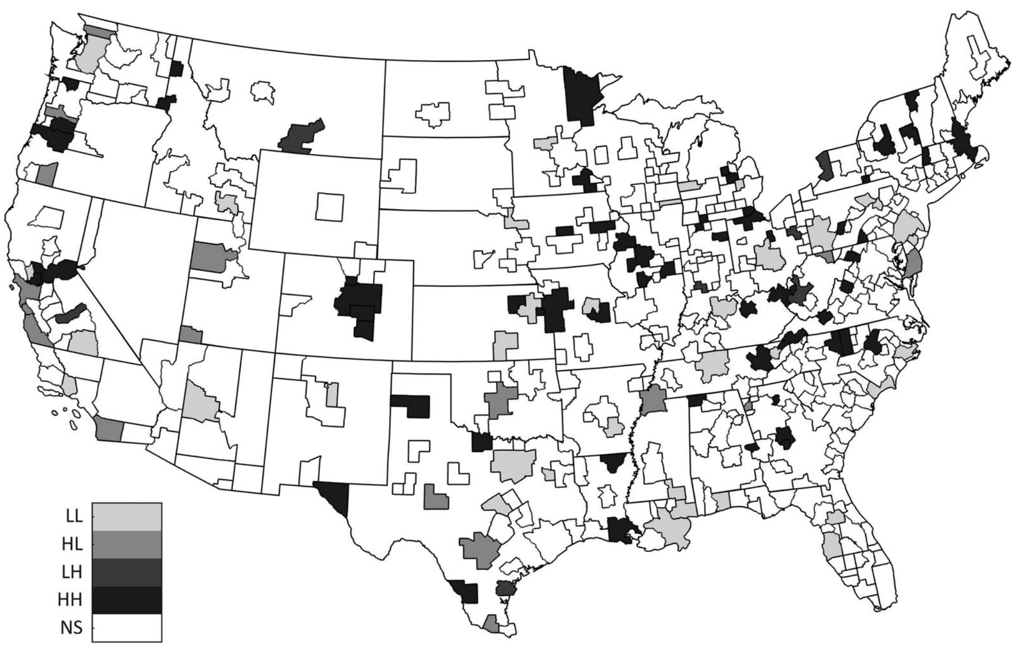
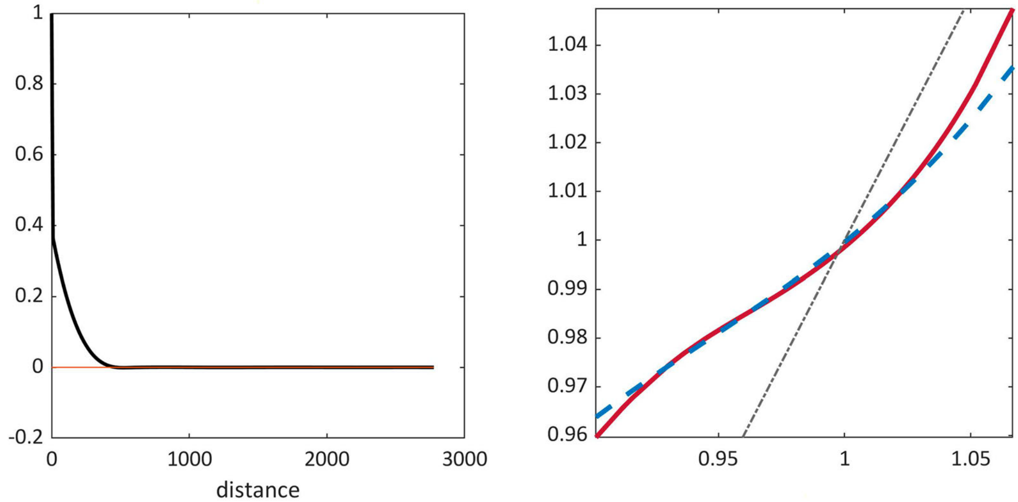


Figure 6. Moran significance map (MSAs). Note: The map shows statistically significant ($p < 0.10$) local Moran's I-statistics using a 10% nearest-neighbours, row-standardized W matrix.



(a) spatial correlogram (SNP) (b) mean function estimates

Figure 7. Spatial correlogram and mean function estimates (MSAs). Note: In the left panel, the number of knots in the estimate of the spline correlogram is chosen using cross-validation. In the right panel, the continuous line represents the SNP estimate, the dashed line represents the NP estimate, the dash–dotted line represents the 45° line. The (fixed) bandwidth for both estimates is chosen using cross-validation (Hurvich et al., 1998).

We apply this novel version of distribution dynamics approach to reconsider the evidence on convergence dynamics across regional economies in the United States. In particular, we analyse convergence between 1975 and 2008 using data on per capita personal income at two different spatial scales: a broader scale (48 conterminous states, excluding the District of Columbia) and a finer scale (380 MSAs). Results show that both states and MSAs are characterized by a tendency towards convergence. However, important features of this depends on whether the presence of spatial dependence is allowed for: while spatial dependence has the effect of increasing the estimated extent and speed of the convergence process, the opposite is true in the case of MSAs when speed of convergence is slower. This outcome also highlights the unpredictability of the features of the distortion due to neglecting the spatial structure brought into the residuals of the mean function estimate; in our view, this reinforces the usefulness of the SNP procedure.

Possible development of this work, currently among our future research lines, is the extension to a panel set-up that should include a dynamic evolution of the spatial dependence.

ACKNOWLEDGMENTS

We are grateful to the referees and editor for very insightful comments. We also thank Roberto Casarin, Davide Fiaschi, Christopher Parmeter and the participants at the ERSA, SEA and SNDE conferences for useful suggestions. All errors and infelicities of style are our own responsibility.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

NOTES

[1] For discussions about the merits of the approach relative to alternative ones and, in particular, to β-convergence, see, among others Durlauf et al. (2005), Durlauf and Quah (1999), Islam (2003) and Magrini (2004, 2009).
[2] Hyndman et al. (1996) show this by adopting a product kernel for the joint density function $\hat{f}_{t,t+s}(y, y')$.
[3] For further details about the properties of the kernel function, see, for example, Azzalini and Bowman (1997).
[4] Within the literature on parametric spatial regression, this set-up is commonly referred to as the spatial error model (Anselin, 1988).
[5] As pointed out by Anselin (2003), an important advantage of the direct representation is that it allows one to overcome forms of non-stationarity in variance, possibly affecting spatial autoregressive models.
[6] This is ensured by selecting a functional form $\rho(\cdot)$, such that $\rho(d_{ij}, \phi) = 1$ for $d_{ij} = 0$.
[7] In what follows, $\tilde{\rho}(.)$ depends only on distance; being a non-parametric estimator, it does not assume any specific functional form for the relation.
[8] Silverman (1984) points out that the smoothing spline is essentially a local kernel average with a variable bandwidth.
[9] The code was written in Matlab (Release R2019b).
[10] To guarantee the required degree of undersmoothing, the bandwidth in the pilot estimate of the SNP estimator is $h = N^{(-1/10)}g$, where g is an optimal bandwidth obtained via either criteria.
[11] Operatively, the estimate of the spatial autocorrelation function is obtained through two subsequent smoothings, and the corresponding smoothing parameters are chosen by generalized cross-validation minimization.
[12] Appendix A1 in the supplemental data online presents an analogous set of tables using a direct plug-in bandwidth.
[13] It is a matrix norm defined as the square root of the sum of the absolute squares of the matrix elements.
[14] For both sets, data have been downloaded from the Bureau of Economic Analysis's website. Data and Matlab codes to replicate the analysis are available at https://sites.google.com/a/unive.it/smagrini/home.
[15] Based on US business cycle dating provided by the National Bureau of Economic Research (NBER).
[16] When using SNP, the spatial correlogram is estimated using a matrix of orthodromic distances between state capitals or MSAs.
[17] In this type of plots, the 45° line highlights persistence properties.

ORCID

Margherita Gerolimetto http://orcid.org/0000-0002-7604-2667
Stefano Magrini http://orcid.org/0000-0001-6999-4558

REFERENCES

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Springer.

Anselin, L. (2003). Spatial econometrics. In B. H. Baltagi (Ed.), *A companion to theoretical econometrics, volume 1*, pp. 310–330. Elsevier, Amsterdam Blackwell.

Azzalini, A., & Bowman, A. W. (1997). *Applied smoothing techniques for data analysis*. Clarendon Press.

Bartels, C. P., & Ketellapper, R. (1979). *Exploratory and explanatory analysis of spatial data*. Martinus Nijhoff.

Basile, R. (2010). Intra-distribution dynamics of regional per-capita income in Europe: Evidence from alternative conditional density estimators. *Statistica*, *70*, 3–22.

Bennett, R. (1979). *Spatial time series*. Pion.

Bjørnstad, O. N., & Falck, W. (2001). Nonparametric spatial covariance functions: Estimation and testing. *Environmental and Ecological Statistics*, *8*(1), 53–70. doi:10.1023/A:1009601932481

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, *74*(368), 829–836. doi:10.1080/01621459.1979.10481038

Corrado, L., & Fingleton, B. (2016). *The W matrix in network and spatial econometrics: Issues relating to specification and estimation*. CEIS Working Paper No. 369, Center for Economic and International Studies, Rome.

Durlauf, S. N., Johnson, P. A., & Temple, J. R. W. (2005). Growth econometrics. In P. Aghion & S. N. Durlauf (Eds.), *Handbook of economic growth, volume 1*, pp. 555–677. Elsevier.

Durlauf, S. N., & Quah, D. T. (1999). The new empirics of economic growth. In J. Taylor & M. Woodford (Eds.), *Handbook of macroeconomics* (pp. 235–308). Elsevier.

Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications*. Chapman & Hall.

Fischer, M. M., & Stumpner, P. (2008). Income distribution dynamics and cross-region convergence in Europe. Spatial filtering and novel stochastic kernel representations. *Journal of Geographical Systems*, *10*(2), 109–139. doi:10.1007/s10109-008-0060-x

Gerolimetto, M., & Magrini, S. (2017). A novel look at long-run convergence dynamics in the United States. *International Regional Science Review*, *40*(3), 241–269. doi:10.1177/0160017614550081

Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, *24*(3), 189–206. doi:10.1111/j.1538-4632.1992.tb00261.x

Green, P. J., & Silverman, B. W. (1994). *Nonparametric regression and generalized linear models: A roughness penalty approach*. Chapman & Hall.

Hall, P., Fisher, N. I., & Hoffman, B. (1994). On the nonparametric estimation of covariance functions. *Annals of Statistics*, *22*, 2115–2134. doi:10.1214/aos/1176325774

Hall, P., & Patil, P. (1994). Properties of nonparametric estimators of autocovariance for stationary random fields. *Probability Theory and Related Fields*, *99*(3), 399–424. doi:10.1007/BF01199899

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer.

Hodrick, R., & Prescott, E. C. (1997). Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, *29*(1), 1–16. doi:10.2307/2953682

Hurvich, C. M., Simonoff, J. S., & Tsai, C.-L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *60*(2), 271–293. doi:10.1111/1467-9868.00125

Hyndman, R. J., Bashtannyk, D. M., & Grunwald, G. K. (1996). Estimating and visualizing conditional densities. *Journal of Computational and Graphical Statistics*, *5*, 315–336. doi:10.1080/10618600.1996.10474715

Hyndman, R. J., & Wand, M. P. (1997). Nonparametric autocovariance function estimation. *Australian Journal of Statistics*, *39*(3), 313–324. doi:10.1111/j.1467-842X.1997.tb00694.x

Islam, N. (2003). What have we learnt from the convergence debate? *Journal of Economic Surveys*, *17*(3), 309–362. doi:10.1111/1467-6419.00197

Johnson, P. A. (2005). A continuous state space approach to "convergence by parts". *Economics Letters*, *86*(3), 317–321. doi:10.1016/j.econlet.2004.06.023

LeSage, J. P., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC Press.

Lin, X., & Carroll, R. J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association*, *95*(450), 520–534. doi:10.1080/01621459.2000.10474229

Magrini, S. (2004). Regional (di)convergence. In J. V. Henderson & J. F. Thisse (Eds.), *Handbook of regional and urban economics, volume 4*, pp. 2741–2796. Elsevier.

Magrini, S. (2009). Why should We analyse convergence using the distribution dynamics approach? *Scienze Regionali – Italian Journal of Regional Science*, *8*, 5–34. doi:10.3280/SCRE2009-001001

Magrini, S., Gerolimetto, M., & Duran, H. E. (2015). Regional convergence and aggregate business cycle in the United States. *Regional Studies*, *49*(2), 251–272. doi:10.1080/00343404.2013.766319

Martins-Filho, C., & Yao, F. (2009). Nonparametric regression estimation with general parametric error covariance. *Journal of Multivariate Analysis*, *100*(3), 309–333. doi:10.1016/j.jmva.2008.04.013

Maza, A., Hierro, M., & Villaverde, J. (2010). Measuring intra-distribution income dynamics: An application to the European regions. *The Annals of Regional Science*, *45*(2), 313–329. doi:10.1007/s00168-009-0304-9

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications*, *9*(1), 141–142. doi:10.1137/1109020

Paelinck, J., & Klaassen, L. (1979). *Spatial econometrics*. Saxon House.

Quah, D. T. (1993a). Empirical cross-section dynamics in economic growth. *European Economic Review*, *37*(2–3), 426–434. doi:10.1016/0014-2921(93)90031-5

Quah, D. T. (1993b). Galton's fallacy and tests of the convergence hypothesis. *The Scandinavian Journal of Economics*, *95*(4), 427–443. doi:10.2307/3440905

Quah, D. T. (1996a). Empirics for economic growth and convergence. *European Economic Review*, *40*(6), 1353–1375. doi:10.1016/0014-2921(95)00051-8

Quah, D. T. (1996b). Convergence empirics across economies with (some) capital mobility. *Journal of Economic Growth*, *1*(1), 95–124. doi:10.1007/BF00163344

Quah, D. T. (1997). Empirics for growth and distribution: Stratification, polarization, and convergence clubs. *Journal of Economic Growth*, *2*(1), 27–59. doi:10.1023/A:1009781613339

Ravn, M. O., & Uhlig, H. (2002). On adjusting the Hodrick–Prescott filter for the frequency of observations. *The Review of Economics and Statistics*, *84*(2), 371–376. doi:10.1162/003465302317411604

Robinson, P. M. (1987). Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. *Econometrica*, *55*(4), 875–891. doi:10.2307/1911033

Robinson, P. M. (2008). Developments in the analysis of spatial data. *Journal of the Japanese Statistical Society*, *38*(1), 87–96. doi:10.14490/jjss.38.87

Robinson, P. M. (2011). Asymptotic theory for nonparametric regression with spatial data. *Journal of Econometrics*, *165*(1), 5–19. doi:10.1016/j.jeconom.2011.05.002

Ruckstuhl, A. F., Welsh, A. H., & Carroll, R. J. (2000). Nonparametric function estimation of the relationship between two repeatedly measured variables. *Statistica Sinica*, *10*, 51–71.

Shorrocks, A. F. (1978). The measurement of mobility. *Econometrica*, *46*(5), 1013–1024. doi:10.2307/1911433

Silverman, B. W. (1984). Spline smoothing: The equivalent variable kernel method. *The Annals of Statistics*, *12*(3), 898–916. doi:10.1214/aos/1176346710

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Chapman and Hall/CRC Press.

Stakhovych, S., & Bijmolt, T. H. A. (2009). Specification of spatial models: A simulation study on weights matrices. *Papers in Regional Science*, *88*(2), 389–408. doi:10.1111/j.1435-5957.2008.00213.x

Wang, N. (2003). Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika*, *90*(1), 43–52. doi:10.1093/biomet/90.1.43

Watson, G. S. (1964). Smooth regression analysis. *Sankhya, Series A*, *26*, 359–372.

Xiao, Z., Linton, O. B., Carroll, R. J., & Mammen, E. (2003). More efficient local polynomial estimation in nonparametric regression with autocorrelated errors. *Journal of the American Statistical Association*, *98*(464), 980–992. doi:10.1198/016214503000000936