Conditional density estimation with spatially dependent data

M. Gerolimetto
Dept. of Statistics
Ca' Foscari University of Venice - Italy
e-mail: margherita.gerolimetto@unive.it

S. Magrini
Dept. of Economics
Ca' Foscari University of Venice - Italy
SSaV (School for advanced Studies in Venice)
e-mail: smagrini@unive.it

Abstract

Estimating conditional distributions for the analysis of convergence is usually done using a standard kernel smoother (Quah, 1997) but this is known to be biased. Hyndman *et al.* (1996) thus suggest a conditional density estimator with a mean function specified by a local polynomial smoother, i.e. one with better bias properties. However, even in this case, the estimated conditional mean might be incorrect when observations are spatially dependent.

In the present paper, we investigate possible developments of the mean bias correction procedure to allow for spatial dependence. In particular, we consider extending usual local linear regression (along the lines of Mathins-Filho and Yao, 2009) by considering a two step method where information on spatial dependence is incorporated in the error covariance matrix, estimated nonparametrically. The finite sample performance of our proposed procedure is then shown via Monte Carlo simulations for various data generating processes.

Finally, it is interesting to note that such an approach would have rather general applicability as it could be employed whenever a local polynomial regression is estimated and the researcher suspects that data might be (spatially) dependent.

Keywords: conditional density estimator, nonparametric smoothing, spatial dependence

JEL Codes: C14, C21

1 Introduction

Distribution dynamics (Quah, 1993 a and b, 1996 a and b, 1997) represents a rather recent 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 $Y_{i,t}$ per capita income of economy i at time t (relative to the group average). Now, denote with $F(Y_t)$ the distribution of Y_t and, assuming it admits a density, indicate this density with $f(Y_t)$. Finally, assume that the dynamics of $F(Y_t)$, or equivalently of $f(Y_t)$, can be modelled as a first order process. As a result, the density prevailing at time $t+s$ is given by

$$f\left(Y_{t+s}\right) = \int_{-\infty}^{\infty} f\left(Y_{t+s}|Y_t\right)\ f\left(Y_t\right)\ dY \tag{1}$$

where the stochastic kernel $f(Y_{t+s}|Y_t)$ maps the density at time t into the density at time $t+s$. This element is the corner-stone of the approach as its (nonparametric) estimate provides information both on the change in the external shape of the distribution and, more importantly, on the movement of the economies from one part of the distribution to another between time t and time $t+s$.

Effectively, the stochastic kernel in equation (1) is a conditional density function, a non-paramentric estimate of which can be obtained by dividing the estimate of the joint probability density function $f(Y_t, Y_{t+s})$ by the estimate of the marginal probability density function $f(Y_t)$:

$$\hat{f}\left(Y_{t+s}|Y_t\right) = \frac{\hat{f}\left(Y_t, Y_{t+s}\right)}{\hat{f}\left(Y_t\right)} \tag{2}$$

The most commonly adopted method to obtain such an estimate is the kernel density estimator. However, Hyndman *et al.* (1996) suggest that this popular estimator might have poor bias properties. To clarify this, consider the following general setting: assume X and Y are

[1]For discussions about the merits of the approach relative to alternative ones and, in particular, to β-convergence see, among others, Durlauf and Quah, 1999; Islam, 2003; Magrini, 2004 and 2009; Durlauf et al., 2005.

an explanatory and a dependent variable, respectively; also, denote by $\{(X_1, Y_1),\ (X_2, Y_2),\ \ldots,\ (X_n, Y_n)\}$ a sample of length n, and by $\{(x_1, y_1),\ (x_2, y_2),\ \ldots,\ (x_n, y_n)\}$ the observations. Indicate the conditional mean with $m(x) = E(Y|X = x)$ so that:

$$Y_j|(X_j = x_j) = m(x_j) + \epsilon_j \quad j = 1, ..., n \tag{3}$$

where the ϵ_j are zero mean and independent, although not necessarily identically distributed.

The kernel estimator of the conditional density of Y conditional on $X = x$ is:

$$\hat{f}(y|x) = \sum_{j=1}^{n} w_j(x) K_b\left(y - Y_j\right) \tag{4}$$

where

$$w_j(x) = \frac{K_a\left(x - X_j\right)}{\sum_{j=1}^{n} K_a\left(x - X_j\right)} \tag{5}$$

a and b are bandwidth parameters controlling the smoothness in, respectively, the X dimension and the Y dimension, $K_b(u) = b^{-1}K(\frac{u}{b})$ is a scaled kernel function and $K(.)$ is assumed to be a real value, integrable and non negative even function[2]. In addition, the mean of the conditional density estimator in (4) provides an estimator of the conditional mean function $m(x)$:

$$\hat{m}(x) = \int Y \hat{f}(y|x) dy = \sum_{j=1}^{n} w_j(x) Y_j. \tag{6}$$

As highlighted by Hyndman *et al.* (1996), note that the estimator in (6) is equivalent to the local constant (or Nadaraya-Watson) regression estimator. This is known to be biased on the boundary of the X space and also in the interior, especially when the mean function is characterized by an evident curvature or simply the scatter plot of the design points is irregular. Calling this bias in the estimated mean as the *mean-bias* of a conditional density estimators, it follows that the kernel estimator of a conditional density shown in (4) can have a large *mean-bias*.

[2]For further details about the properties of the kernel function, see, for example, Azzalini and Bowman (1997).

As an alternative, Hyndman *et al.* (1996) then propose a new class of conditional density estimators, defined as:

$$\hat{f}^*(y|x) = \sum_{j=1}^{n} w_j(x) K_b \left(y - Y_j^*(X) \right) \tag{7}$$

where $Y_j^*(x) = \hat{m}(x) + e_j - \sum_{i=1}^{n} w_i(x) e_i$, and $E_i = Y_i - \hat{m}(x_i)$, $i = 1, ..., n$.

By construction, the mean-bias of the estimator in (7) is equal to a previously estimated $\hat{m}(x)$. Clearly, this means that when $\hat{m}(x)$ is the Nadaraya-Watson smoother, the estimator reverts to the traditional kernel density estimator in (4). More importantly, it also suggests that a lower mean-bias con be obtained by employing a smoother with better bias properties than kernel smoothing. One such smoother is, for instance, the local linear estimator (Loader, 1999):

$$\hat{m}(x) = \frac{\sum_{j=1}^{n} K_a(x - X_j) Y_j}{\sum_{j=1}^{n} K_a(x - X_j)} + (x - \bar{X}_w) \frac{\sum_{j=1}^{n} K_a(x - X_j)(X_j - \bar{X}_w) Y_j}{\sum_{j=1}^{n} K_a(x - X_j)(X_j - \bar{X}_w)^2} \tag{8}$$

where

$$\bar{X}_w = \frac{\sum_{j=1}^{n} K_a(x - X_j) X_j}{\sum_{j=1}^{n} K_a(x - X_j)}$$

It is important to emphasise that the asymptotic properties of the smoothers employed to estimate $m(x)$ are based on the assumption that the error terms of the nonparametric regression (3) are zero mean and uncorrelated variables. However, in empirical analyses of cross-sectional convergence it is highly unlikely that the data comply with this hypothesis. Instead, data are often characterized by spatial dependence, i.e. a form of dependence that varies with the relative location or distance between observations.

Within the parametric context, a possible, rather general, model for spatially dependent data is the Spatial Durbin Model (Anselin, 1988):

$$Y = \rho W Y + \beta X - \lambda W \beta X + \epsilon \tag{9}$$

where X and Y are $n \times 1$ vectors, ϵ is a $n \times 1$ vector of innovations, $-1 < \rho < 1$ and $-1 < \lambda < 1$. W is a $n \times n$ spatial weights matrix with non negative w_{ij} elements and zeros on the diagonal. A

special case of the above model is the Spatial Lag Model, derived from (9) when the constraint $\lambda = 0$ holds:

$$Y = \rho WY + \beta X + \epsilon \tag{10}$$

Another commonly considered case is the Spatial Error Model, which can be derived from model (9) when the constraint becomes $\rho = \lambda$:

$$\begin{array}{c} Y = \beta X + u \\ u = \lambda W u + \epsilon \end{array} \tag{11}$$

Asymptotic properties of various (parametric) estimates of ρ, λ and β have been studied. Anselin (1988) shows the inconsistency of the Least Squares estimates of λ and β. Kelejan and Prucha (1998, 1999) show that other estimates, such as Instrumental Variables and Gaussian Maximum Likelihood, are consistent and asymptotically normal.

For convergence analyses adopting the distribution approach, model (11) seems of particular interest as the $\rho = \lambda$ constraint is likely to hold given that X and Y represent the same phenomenon observed at two different points in time. However, as recalled above, the central element of the approach, the stochastic kernel in equation (1), is estimated nonparametrically and parametric models like (11) do not appear to be appropriate. Consequently, we consider a direct extension[3] of model (11) to the nonparametric framework:

$$\begin{array}{c} Y = m(X) + u \\ u = \lambda W u + \epsilon \end{array} \tag{12}$$

As for the estimation of a nonparametric regression when data exhibit spatial dependence, Robinson (2008, 2009) derives consistency and asymptotic distribution theory for the local constant regression estimator for various kinds of spatial data, including those generated by model (12) for u on a lattice of arbitrary dimension. Other authors (for example, Xiao *et al.*, 2003; Lin and Carroll, 2000; Ruckstuhl *et al.*, 2000; Wang, 2003) study possible extensions of the nonparametric regression to a non i.i.d. errors setting, where errors can be correlated

[3]It is also possible to generalize further; for example, one could have $Y = g(WY, m(X)) + u$, where $u = \lambda W u + \epsilon$

and heteroschedastic. In all cases, however, a parametric structure for the dependence must be assumed beforehand. In the present paper, we propose a two step procedure for nonparametric regression with spatially dependent data that does not require *a priori* parametric assumptions on spatial dependence as information on its structure is instead drawn from a nonparametric estimate of the errors spatial covariance matrix.

The structure of the paper is as follows. In the second Section we present our procedure, after illustrating three issues connected to our proposal: *i)* nonparametric regression with spatial dependent error terms, *ii)* nonparametric estimation of the spatial covariance matrix and *iii)* the bandwidth selection using spatial dependence. In the third Section we present the finite sample experiment we realized and the results.

2 A new procedure for nonparametric regression with spatially dependent errors

2.1 Non parametric regression with dependent data

In a very recent paper, Martins-Filho and Yao (2009) (hereafter MFY) propose a two step procedure that incorporates information about the (spatial) dependence structure in the error terms. For the nonparametric regression (12), the authors' idea is, firstly, to obtain a pilot $\hat{m}(X)$ nonparametric estimate of function $m(X)$, then estimate the spatial covariance matrix V of the residuals $\hat{u} = Y - \hat{m}(X)$, and finally obtain $\check{m}(X)$ by running a modified regression, where the spatial covariance matrix V is opportunely included in order for the error terms to be spherical. MFY demonstrate the asymptotic properties of their two step procedure under rather general conditions on the covariance structure of the errors. In more detail, for model (12) they assume u to be a vector of error terms such that $E(u_i) = 0$, $\forall i = 1, ..., n$, and $E(u_i, u_j) = \omega_{ij}(\theta_0)$, $\theta_0 \in R^p$, $p < \infty$. The estimate of the regression function is then obtained using the local linear estimator in (8). In order to obtain a model with spherical errors, MFY

define $z = P^{-1}(\theta_0)Y + (I - P^{-1}(\theta_0))\hat{m}(X)$, so that the modified model becomes

$$z = m(X) + P^{-1}(\theta_0)u = m(X) + \epsilon \tag{13}$$

which is shown (MFY, 2009; Theorem 3) to be an improvement over the traditional local linear estimator having bias of the same order, but an asymptotic distribution with strictly smaller variance. These properties continue to hold if matrix V is not known and has to be replaced by an estimate $\hat{V}$, but the latter must be consistent in order to guarantee the asymptotic properties to hold also for the feasible version of the estimator (MFY, 2009; Theorem 3).

2.2 Nonparametric estimation of spatial dependence

In a spatial error model (both in its parametric and nonparametric version), the error term u is assumed to be a spatial autoregression (SAR), whose specification is $u = \lambda Uu + \epsilon$, where ϵ are spatial white noise (i.e. $E(\epsilon) = 0$ and $E(\epsilon_i, \epsilon_j) = \sigma^2$, for $i = j$ and $E(\epsilon_i, \epsilon_j) = 0$, for $i \neq j$). Under these circumstances, the spatial covariance matrix of u is

$$V = \sigma^2 \left[(I - \lambda W)'(I - \lambda W)\right]^{-1} \tag{14}$$

The elements of the covariance matrix V can be also represented directly (for example, Arbia, 2006) as some function of the distance between pairs of sites. Hence, a spatial covariance function can be defined as follows:

$$\gamma(s_i, sj) = \sigma^2 f(d_{ij}, \phi) \tag{15}$$

where d_{ij} is the distance between sites i,j and $f(.)$ is a decaying function such that $\frac{\partial f}{\partial d_{ij}} < 0$, $|f(d_{ij}, \phi| \leq 1$ and ϕ is an appropriate vector of parameters. Following this approach, the full matrix of spatial covariance is $V = \sigma^2\Omega(d_{ij}, \phi)$, which must be positive definite and with elements $w_{ij} \in \Omega$, such that $|w_{ij}| \leq 1$ for each i and j[4]. The direct representation of the spatial

[4]Now every element of the covariance matrix is modeled so that the problem of nonstationary of the SAR model is overcome.

covariance matrix has the obvious advantage of being parsimonious; in the present context, it has the even more important advantage of allowing a nonparametric estimation of spatial dependence instead of requiring a parametric estimate of λ to be plugged-in in (14) and the specification of a matrix of spatial weights.

Given our aim of estimating nonparametrically the spatial covariance of the error terms of a nonparametric regression, we then follow Bjørnstad and Falk (2001) (hereafter, BF), who develop a continuous nonparametric positive semidefinite estimator of the covariance function, called spline correlogram, building on the work of Hall and coworkers (Hall and Patil, 1994; Hall *et al.*, 1994).

Hall and Patil (1994) propose a kernel estimator of the spatial autocorrelation function $\rho(s_i, s_j)$. This estimator, based on the sample correlation

$$\hat{\rho}_{ij} = \frac{(z_i - \bar{z})(z_j - \bar{z})}{1/n \sum_{l=1}^{n} (z_l - \bar{z})^2} \tag{16}$$

where $\bar{z} = 1/n \sum_{l=1}^{n} z_l$ is the sample mean, is as follows:

$$\tilde{\rho}(d) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} K(d_{ij}/h)(\hat{\rho}_{ij})}{\sum_{i=1}^{n} \sum_{j=1}^{n} K(d_{ij}/h)} \tag{17}$$

where K is a kernel function and h is a bandwidth. The authors also showed the asymptotic properties of the kernel estimator in (17) for certain values of h.

BF (2001) propose several improvements over the estimator in (17). Firstly, they use a cubic B-spline as an equivalent kernel smoother because it seems to adapt better to irregularly spaced data and produces a consistent estimate of the covariance function (Hyndman and Wand, 1997). The asymptotic kernel function for the cubic B-spline[5] is (see Green and Silverman (1994) for more details):

$$K(u) = \frac{1}{2} exp\left(-\frac{|u|}{\sqrt{2}}\right) sin\left(-\frac{|u|}{\sqrt{2}} + \frac{\pi}{4}\right) \tag{18}$$

Secondly, the estimator $\tilde{\rho}$ must be not only consistent, but also positive semidefinite. This means (Bochner's theorem) that the Fourier transformed function has to be strictly non-negative and

[5]The degree of smoothing is represented by the equivalent degrees of freedom, corresponding to the number of effective parameters that are involved in the fit.

this is not guaranteed by the estimator in equation (17). Hence, BF use the Fourier-filter methods by Hall *et al.* (1994): firstly the Fourier transform of $\tilde{\rho}_i j$ is obtained, then all negative excursions of the transformed function are set to zero and finally, by backtransformation, a non parametric positive semidefinite estimate of the spatial covariance is obtained.

The end result is a nonparametric estimator of the spatial correlation which BF call spline correlogram. The assumptions[6] on which this estimator is based are: *i)* data are generated from a second order stationary random field, i.e. the expectation and covariance function do not change through space; *ii)* the field is isotropic, i.e. the covariance only depends on distance and not direction; *iii)* data are Gaussian. Under these circumstances, the spatial correlation function is $\rho(s_i, s_j) = 1/\sigma^2\gamma(s_i, s_j)$.

2.3 Choosing the bandwidth parameter through spatial statistics

Ellner and Seifu (2002), herafter ES, propose a method (RSA) to select the bandwidth h in a nonparametric regression, based on residuals spatial dependence. The basic idea here is that, in case of unfitted structure, two nearby residuals tend to show a greater degree of similarity (on average) than two randomly chosen residuals. So, when the residuals exhibit spatial correlation in the space defined by the independent variables, this is a signal that the model is not capturing all the structure in the data and a different bandwidth has to be preferred.

To quantify spatial dependence, in direct analogy with the use of the familiar Durbin Watson test in time series data, ES make use of Moran's I statistic:

$$I = \frac{n \sum_{i \neq j=1}^{n} w_{ij}(e_i - \bar{e})(e_j - \bar{e})}{\sum_{i \neq j=1}^{n} w_{ij} \sum_{i=1}^{n}(e_i - \bar{e})^2} \tag{19}$$

This test statistics is used to test for the null hypothesis of no spatial dependence in the residuals. The significance of non zero values of I can be established using the asymptotic normal distribution. However, especially when the data do not comply with all the hypotheses, consistency of the tests is proved when the reference distribution of the I statistics obtained

[6]These assumptions coincide with the occurrence of data generated by a SAR model when each site has an identical number of neighbours. The latter condition guarantees second order stationarity.

under random permutation of the residuals. However, for moderate to large sample sizes, the sample values of I will be based on a large number of pairs and so the critical values for testing the null hypothesis can be often approximated using a normal distribution with mean and standard error of $\hat{I}$ under random permutation of the residuals.

From an operational point of view, an appropriate value for the bandwidth can therefore be chosen as follows: *i)* select a grid of possible values for the bandwidth h; *ii)* compute Moran's I on the residuals from each of possible values of h; *iii)* plot these as a function of h and then choose the value of h such that I matches its expected value. In other words, if a good specification for the model is reached, a reasonable value for h is the one that guarantees that the null hypothesis of no spatial dependence in the residuals is accepted.

2.4 A new procedure

In this paper we to propose a new procedure to estimate nonparametrically the model (12) in which errors are spatially dependent. To do this, we develop a two-stage procedure that moves along the lines of MFY (2009), but with two main differences. Firstly, we estimate the covariance matrix nonparametrically, rather than parametrically, using the spline correlogram proposed by BF (2001). This is done in order the procedure to be more general, allowing it to disentagle itself from many parametric requirements. Secondly, the information on spatial dependence is exploited also in the final choice of the bandwidth parameter.

The following is a schematic description of our procedure:

1. *Pilot fit*: estimate $m(X)$ with a local polynomial smoother, where h is in fact a nearest neighbour smoothing parameter such that the width of each smoothing neighborhood covers 30% of the data. As for the degree of the polynomial, we consider $p = 0, 1$. Thus the local constant and the local linear estimators are effectively used. The output is $\hat{u} = Y - \hat{m}(X)$.

2. *Non parametric covariance matrix estimation*: using BF (2001) spline correlogram, obtain $\hat{V}$, the estimated spatial covariance matrix of $\hat{u}$.

3. *Final fit*: feed the procedure with the information obtained from the estimate of the spatial covariance matrix $\hat{V}$ by running a modified regression where Y is replaced by $Z = \hat{m}(X) + L^{-1}\hat{u}$ and L is obtained by taking the Cholevsky decomposition of $\hat{V}$. The nonparametric estimate $\check{m}$ resulting from this second fit is done by choosing the bandwidth parameter with a modified version of the RSA criterion[7] suggested by ES (2002). In particular, the adaptations we introduced to the original RSA criterion are:

   - space is geographical space rather than the variables' space;
   - inference on Moran's I is based on random permutations of the residuals;
   - the spatial weight matrix for conducting the test is constructed as follows: from the spline-based estimate of spatial correlation function $\hat{V}$, we determine the distance d_{zc} at which correlation becomes zero; next, we define locations i and j to be neighbours, and set $w_{ij} = 1$, if their distance d_{ij} is smaller than d_{zc}; otherwise, we set $w_{ij} = 0$.

3 Monte Carlo study

In this Section we describe our Monte Carlo experiment that aims at showing, via simulations, the finite sample performance of our procedure in comparison with other estimation methods, both parametric and nonparametric, that do not take into account the presence of spatial dependence. The code has been written in Matlab (Matlab 7.7.0, R2008b). The comparison is done in order to show the effective improvement in regression estimation results when spatial dependence is not neglected. The performance of the estimation methods is expressed in terms of percentage of rejections of the Moran's I test on the residuals as well as in terms of the Monte Carlo mean and standard deviations of the Moran's I index computed on the residuals.

The Monte Carlo experiment consists of two parts. Firstly, the performance of all estimators

[7]For computational reasons, the choice of h is done by minimizing the p-value instead of solving the equation $I(h) = 0$ as originally suggested by ES

has been compared in case of data generated from linear systems:

$$\begin{aligned} Y &= X + u \\ u &= \lambda W u + \epsilon \end{aligned} \tag{20}$$

where $X \sim N(100, 1)$ and $\epsilon \sim N(0, 1)$. The distance matrix is based on euclidean distances computed over spatial coordinates generated as independent draws from $U(0, 1000)$. The spatial weights matrix is obtained as a k-neighbours contiguity matrix, where k is 20% of the number of observations.

Secondly, the performance of all estimators has been compared in case of data generated from non linear systems:

$$\begin{aligned} Y &= m(X) + u \\ u &= \lambda W u + \epsilon \end{aligned} \tag{21}$$

where $m(X) = sin(2X) + 2e^{-2X^2}$, X, ϵ and W are generated as described above.

Various values of parameter λ have been considered, in particular $\lambda = 0, 0.2, 0.4, 0.8$, in order to establish the performance for situations where spatial dependence is absent, weak, medium and strong. The simulated data set length is $N = 50, 100, 200$ and the number of Monte Carlo simulations is 2000. The Moran's I statistics significance ($\alpha = 0.05$) has been established using 999 random permutations.

The estimation methods we used are: our procedure implemented both with a local constant and a local linear estimator (in the tables SNP0 and SNP1); traditional local constant and local linear (in the tables NP0 and NP1); Ordinary Least Squares (in the tables OLS).

In tables 1-12 (at the end of the paper) we present the results of our simulations. Specifically, we report in tables 1, 3 and 5, the percentages of rejections of the null hypothesis of absence of spatial dependence in the residuals for linear systems; in tables 7, 9 and 11, we report the corresponding percentages for non linear systems. The idea is that the better the estimate, the more dependence is captured by $\hat{m}(X)$, the less the Moran's I null hypothesis is rejected.

In tables 2, 4, 6, 8, 10, 12 we report the corresponding Monte Carlo mean and standard deviation of Moran's I index. The smaller the value of I, the better the diagnostics on the residuals because the lower is the degree of spatial dependence left in the residuals.

As we can see from the tables, nonparametric methods (NP0, NP1, SNP0, SNP1) work in general better than OLS. They tend to reject the null hypothesis of spatial independence in the residuals less often as well as they exhibit smaller values of Moran's I index. This is confirmed both for linear and non linear systems, although this less evident in the linear case as, here, the OLS represents a rival very difficult to beat.

Among nonparametric methods, the SNP procedure visibly outperforms the traditional polynomial regression; for both ploynomial degrees, SNP appears able to detect the spatial structure in the system and consequently to estimate $m(X)$ leaving no spatial dependence in the residuals.

As expected, best preformances of the SNP procedure are obtained for $0 \leq \lambda \leq 0.4$. This is particularly evident in the non linear case where the percentage of rejection does not exceed 0.05 even for $\lambda = 0.4$.

Obvioulsy, the highest percentages of rejection are found when $\lambda = 0.8$. However, it must be emphasised that the percentages achieved by the SNP procedure are considerably lower than those obtaiened by the other methods.

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrect: Kluwer.

Arbia, G. (2006) *Spatial Econometrics: Statistical Foundations and Applications to Regional Analysis*. Berlin: Springer-Verlag.

Azzalini, A. and Bowman A.W. (1997) *Applied Smoothing Techniques for Data Analysis*. Clarendon Press, Oxford.

Bjørnstad, O.N. and Falck, W. (2001). Nonparametric Spatial Covariance Functions: Estimation and Testing. *Enviromental and Ecological Statistics*, **8**, 53–70.

Cliff, A.D. and Ord, J.K. (1973). *Spatial Autocorrelation*. London: Pion.

Durlauf, S.N., Johnson, P.A. and Temple, J.R.W. (2005). Growth Econometrics. In *Handbook of Economic Growth, Volume 1*, Aghion, P. and Durlauf, S.N. (Eds.), Amsterdam: Elsevier.

Durlauf, S.N. and Quah, D.T. (1999). The New Empirics of Economic Growth. In *Handbook of Macroeconomics*, Taylor, J. and Woodford, M. (Eds.), Amsterdam: Elsevier.

Ellner, S.P, O.N. and Seifu, Y. (2002). Using Spatial Statistics to Select Model Complexity. *Journal of Computational and Graphical Statistics*, **11**, 348–369.

Green, P.J. and Silverman B.W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. London: Chapman and Hall.

Hall, P. and Patil, P. (1994). Properties of Nonparametric Estimators of Autocovariance for Stationary Random Fields. *Probability Theory and Related Fields*, **99**, 399–424.

Hall, P., Fisher, N.I and Hoffman, B. (1994). On the Nonparametric Estimation of Covariance Functions. *Annals of Statistics*, **22**, 2115–2134.

Hyndman, R.J., Bashtannyk, D.M. and Grunwald, G.K. (1996). Estimating and Visualizing Conditional Densities. *Journal of Computational and Graphical Statistics*, **5**, 315–336.

Hyndman, R.J. and Wand M.P. (1997). Nonparametric Autocovariance Function Estimation. *Australian Journal of Statistics*, **39**, 313–324.

Islam N. (2003). What Have We Learnt from the Convergence Debate. *Journal of Economic Surveys*, **17**, 309–362.

Kelejian, H.H. and Prucha, I.R. (1998). A Generalized Spatial two-stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. *Journal of Real Estate and Economics*, **17**, 99–121.

Kelejian, H.H. and Prucha, I.R. (1999). A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic Review*, **40**, 509–533.

Lin, X. and Carroll, R.J. (2000). Nonparametric Function Estimator for Clustered Data when the Predictor is Measured with/without Error. *Journal of the American Statistical Association*, **95**, 520–534.

Loader, C. (1999). *Local Regression and Likelihood*. Berlin: Springer.

Magrini S. (2004). Regional (Di)Convergence. In *Handbook of Regional and Urban Economics, Volume 4*, Henderson, J.V. and Thisse J.F. (Eds.), Amsterdam: Elsevier.

Magrini S. (2009). Why Should We Analyse Convergence Using the Distribution Dynamics Approach? *Scienze Regionali - Italian Journal of Regional Science*, **8**, 5–34.

Mardia, K.V. and Watkins A.J. (1989) On Multimodality of the Likelihood for the Spatial Linear Model. *Biometrika*, **76**, 289–295.

Marthins-Filho, C. and Yao, F. (2009). Nonparametric Regression Estimation with General Parametric Error Covariance. *Journal of Multivariate Analysis*, **100**, 309–333.

Quah, D.T. (1993a). Empirical Cross-section Dynamics in Economic Growth. *European Economic Review*, **37**, 426–434.

Quah, D.T. (1993b). Galton's Fallacy and Tests of the Convergence Hypothesis. *Scandinavian Journal of Economics*, **95**, 427–443.

Quah, D.T. (1996a). Empirics for Economic Growth and Convergence. *European Economic Review*, **40**, 1353–1375.

Quah, D.T. (1996b). Convergence Empirics Across Economies with (Some) Capital Mobility. *Journal of Economic Growth*, **1**, 95–124.

Quah, D.T. (1997). Empirics for Growth and Distribution: Stratification, Polarization, and Convergence Clubs. *Journal of Economic Growth*, **2**, 27–59.

Robinson, P.M. (2008). Developments in the Analysis of Spatial Data. *Journal of the Japanese Statistical Society*, **38**, 87–96.

Robinson, P.M. (2009). Asymptotic Theory for Nonparametric Regression with Spatial data. *Journal of Econometrics*, *Forthcoming*.

Rosenblatt, M. (1969). Conditional Probability Density and Regression Estimators. In *Multivariate Analysis II*, Krishnaiah P.R. (Ed.), New York: Academic Press.

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

Wang, N. (2003). Marginal Nonparametric Kernel Regression Accounting for within Subject Correlation. *Biometrika*, **90**, 43–52.

Xiao, Z., Linton, O.B., Carroll, R.J. and Mammen, E. (2003). More Efficient Local Polynomial Estimation in Nonparametric Regression with Autocorrelated Errors. *Journal of the American Statistical Association*, **98**, 980–992.

Table 1: Moran' I empirical probability of rejections - N=50 Linear Systems

λ	OLS	NP0	NP1	SNP0	SNP1
0	0.052	0.051	0.050	0.018	0.008
0.2	0.140	0.123	0.130	0.049	0.024
0.4	0.359	0.320	0.355	0.158	0.040
0.8	0.847	0.802	0.841	0.630	0.153

Table 2: Moran' I Monte Carlo mean (sd in parenthesis) - N=50 Linear Systems

λ	OLS	NP0	NP1	SNP0	SNP1
0	-0.021 (0.053)	-0.022 (0.052)	-0.021 (0.053)	-0.024 (0.041)	-0.022 (0.036)
0.2	0.009 (0.064)	0.004 (0.059)	0.009 (0.064)	-0.006 (0.044)	-0.014 (0.039)
0.4	0.061 (0.095)	0.050 (0.089)	0.058 (0.095)	0.022 (0.063)	-0.004 (0.046)
0.8	0.238 (0.150)	0.207 (0.144)	0.232 (0.146)	0.127 (0.108)	0.015 (0.064)

Table 3: Moran' I empirical probability of rejections - N=100 Linear Systems

λ	OLS	NP0	NP1	SNP0	SNP1
0	0.046	0.043	0.051	0.017	0.009
0.2	0.146	0.137	0.144	0.051	0.023
0.4	0.370	0.297	0.359	0.163	0.024
0.8	0.871	0.828	0.866	0.684	0.099

Table 4: Moran' I Monte Carlo mean (sd in parenthesis) - N=100 Linear Systems

λ	OLS	NP0	NP1	SNP0	SNP1
0	-0.010 (0.026)	-0.011 (0.027)	-0.010 (0.026)	-0.012 (0.021)	-0.012 (0.019)
0.2	0.007 (0.034)	0.004 (0.032)	0.006 (0.034)	-0.001 (0.024)	-0.006 (0.021)
0.4	0.032 (0.047)	0.024 (0.043)	0.031 (0.047)	0.011 (0.032)	-0.004 (0.021)
0.8	0.160 (0.110)	0.136 (0.101)	0.158 (0.108)	0.089 (0.076)	0.003 (0.029)

Table 5: Moran' I empirical probability of rejections - N=200 Linear Systems

λ	OLS	NP0	NP1	SNP0	SNP1
0	0.051	0.041	0.051	0.012	0.009
0.2	0.151	0.124	0.155	0.055	0.024
0.4	0.376	0.324	0.363	0.196	0.051
0.8	0.897	0.847	0.892	0.751	0.221

Table 6: Moran' I Monte Carlo mean (sd in parenthesis) - N=200 Linear Systems

λ	OLS	NP0	NP1	SNP0	SNP1
0	-0.005 (0.013)	-0.005 (0.013)	-0.005 (0.013)	-0.006 (0.010)	-0.006 (0.009)
0.2	0.003 (0.017)	0.002 (0.016)	0.003 (0.017)	-0.001 (0.012)	-0.003 (0.011)
0.4	0.017 (0.025)	0.014 (0.023)	0.017 (0.025)	0.008 (0.017)	0.000 (0.012)
0.8	0.096 (0.072)	0.080 (0.063)	0.095 (0.071)	0.057 (0.049)	0.009 (0.023)

Table 7: Moran' I empirical probability of rejections - N=50 Non Linear Systems

λ	OLS	NP0	NP1	SNP0	SNP1
0	0.054	0.049	0.051	0.016	0.007
0.2	0.115	0.123	0.126	0.048	0.011
0.4	0.239	0.291	0.313	0.152	0.041
0.8	0.804	0.862	0.876	0.737	0.198

Table 8: Moran' I Monte Carlo mean (sd in parenthesis) - N=50 Non Linear Systems

λ	OLS	NP0	NP1	SNP0	SNP1
0	-0.018 (0.051)	-0.017 (0.051)	-0.017 (0.051)	-0.021 (0.042)	-0.022 (0.037)
0.2	0 (0.058)	0.005 (0.059)	0.007 (0.059)	-0.006 (0.043)	-0.014 (0.035)
0.4	0.029 (0.073)	0.042 (0.078)	0.047 (0.081)	0.020 (0.059)	-0.004 (0.045)
0.8	0.235 (0.165)	0.276 (0.175)	0.286 (0.175)	0.186 (0.140)	0.025 (0.077)

Table 9: Moran' I empirical probability of rejections - N=100 Non Linear Systems

λ	OLS	NP0	NP1	SNP0	SNP1
0	0.048	0.051	0.05	0.018	0.003
0.2	0.119	0.123	0.133	0.06	0.014
0.4	0.243	0.309	0.336	0.166	0.027
0.8	0.813	0.861	0.876	0.746	0.098

Table 10: Moran' I Monte Carlo mean (sd in parenthesis) - N=100 Non Linear Systems

λ	OLS	NP0	NP1	SNP0	SNP1
0	-0.010 (0.027)	-0.010 (0.027)	-0.010 (0.027)	-0.011 (0.022)	-0.012 (0.019)
0.2	0.001 (0.032)	0.003 (0.033)	0.005 (0.034)	-0.001 (0.025)	-0.007 (0.020)
0.4	0.019 (0.040)	0.025 (0.043)	0.028 (0.044)	0.013 (0.033)	-0.003 (0.021)
0.8	0.137 (0.108)	0.162 (0.118)	0.171 (0.121)	0.113 (0.093)	0.004 (0.032)

Table 11: Moran' I empirical probability of rejections - N=200 Non Linear Systems

λ	OLS	NP0	NP1	SNP0	SNP1
0	0.048	0.045	0.046	0.018	0.005
0.2	0.12	0.126	0.134	0.068	0.015
0.4	0.271	0.318	0.335	0.215	0.047
0.8	0.836	0.886	0.906	0.82	0.254

Table 12: Moran' I Monte Carlo mean (sd in parenthesis) - N=200 Non Linear Systems

λ	OLS	NP0	NP1	SNP0	SNP1
0	-0.005 (0.013)	-0.005 (0.013)	-0.005 (0.013)	-0.006 (0.011)	-0.0063 (0.009)
0.2	0.000 (0.016)	0.002 (0.017)	0.002 (0.017)	-0.000 (0.013)	-0.006 (0.010)
0.4	0.011 (0.022)	0.014 (0.024)	0.015 (0.024)	0.009 (0.019)	-0.001 (0.012)
0.8	0.078 (0.061)	0.094 (0.069)	0.101 (0.072)	0.074 (0.059)	0.011 (0.027)