

Income convergence and spatial dependence

Margherita Gerolimetto and Stefano Magrini

Abstract Conditional distributions for the analysis of convergence are usually estimated using a standard kernel smoother 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. Consequently, in this paper we study per capita income inequalities among European Functional Regions through a distribution dynamics approach in which the conditional mean is estimated via a procedure that allows for spatial dependence (Gerolimetto and Magrini, 2009).

Key words: Distribution dynamics, Nonparametric smoothing, Spatial dependence

1 Introduction

Distribution dynamics (Quah, 1993, 1996, 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 income¹.

In simple terms, consider a group of n economies and indicate with $y_{i,t}$ the level of 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

Margherita Gerolimetto

Department of Statistics, Ca' Foscari University, Venice, e-mail: margherita.gerolimetto@unive.it

Stefano Magrini

Department of Economics and SSE (Advanced School of Economics in Venice), Ca' Foscari University, Venice e-mail: s.magrini@unive.it

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); Magrini (2004 and 2009); Durlauf *et al.* (2005).

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

$$f(y_{t+s}) = \int_{-\infty}^{\infty} f(y_{t+s}|y_t) f(y_t) dy_t \quad (1)$$

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 $t + s$. Convergence is then analysed directly from the shape of a 3-dimensional plot of the stochastic kernel estimate or from the corresponding contour plot. Additionally, assuming that the process behind (1) is Markov and time homogeneous, convergence dynamics can be analysed by comparing the features of the distribution at time t with the limit of $f(y_{t+s})$ as $s \rightarrow \infty$, the latter being referred to as the stationary (or ergodic) distribution. The stationary distribution represents the density of the cross-sectional income distribution the system would tend to would it be allowed to proceed along its historical path.

Effectively, the stochastic kernel in equation (1) is a conditional density function, a nonparametric estimate of which can be obtained by dividing the estimate of the joint probability density function $\hat{f}(y_t, y_{t+s})$ by the estimate of the marginal probability density function $\hat{f}(y_t)$:

$$\hat{f}(y_{t+s}|y_t) = \frac{\hat{f}(y_t, y_{t+s})}{\hat{f}(y_t)} \quad (2)$$

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, following Hyndman's setting, let Y_t and Y_{t+s} be, respectively, the explanatory and the dependent variable; also, denote by $\{(Y_{t,1}, Y_{t+s,1}), (Y_{t,2}, Y_{t+s,2}), \dots, (Y_{t,n}, Y_{t+s,n})\}$ a sample of length n and by $\{(y_{t,1}, y_{t+s,1}), (y_{t,2}, y_{t+s,2}), \dots, (y_{t,n}, y_{t+s,n})\}$ the observations. Indicate the conditional mean with $m(y_t) = E(Y_{t+s}|Y_t = y_t)$ so that:

$$Y_{t+s,j}|(Y_{t,j} = y_{t,j}) = m(y_{t,j}) + \varepsilon_j \quad j = 1, \dots, n \quad (3)$$

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

The kernel estimator of the conditional density of Y_{t+s} conditional on $Y_t = y_t$ is:

$$\hat{f}(y_{t+s}|y_t) = \sum_{j=1}^n w_j(y_t) K_b(y_{t+s} - Y_{t+s,j}) \quad (4)$$

where

$$w_j(y_t) = \frac{K_a(y_t - Y_{t,j})}{\sum_{j=1}^n K_a(y_t - Y_{t,j})} \quad (5)$$

a and b are bandwidth parameters controlling the smoothness in, respectively, the explicative variable dimension, Y_t , and the dependent variable dimension, Y_{t+s} , $K_b(u) = b^{-1}K(\frac{u}{b})$ is a scaled kernel function and $K(\cdot)$ is assumed to be a real, integrable and non negative even function². In addition, the mean of the conditional density estimator in (4) provides an estimator of the conditional mean function $m(y_t)$:

$$\hat{m}(y_t) = \int y_{t+s} \hat{f}(y_{t+s}|y_t) dy_t = \sum_{j=1}^n w_j(y_t) Y_{t+s,j}. \quad (6)$$

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 explicative variable's space and also in the interior, especially when the mean function is characterized by an evident curvature or 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.

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

$$\hat{f}^*(y_{t+s}|y_t) = \sum_{j=1}^n w_j(y_t) K_b(y_{t+s} - Y_{t+s,j}^*(y_t)) \quad (7)$$

where $Y_{t+s,j}^*(y_t) = \hat{m}(y_t) + e_j - \sum_{i=1}^n w_i(y_t) e_i$, and $E_i = Y_{t+s,i} - \hat{m}(y_{t,i})$, $i = 1, \dots, n$.

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

$$\begin{aligned} \hat{m}(y_t) &= \frac{\sum_{j=1}^n K_a(y_t - Y_{t,j}) Y_{t+s,j}}{\sum_{j=1}^n K_a(y_t - Y_{t,j})} \\ &+ (y_t - \bar{Y}_{tw}) \frac{\sum_{j=1}^n K_a(y_t - Y_{t,j}) (Y_{t,j} - \bar{Y}_{tw}) Y_{t+s,j}}{\sum_{j=1}^n K_a(y_t - Y_{t,j}) (Y_{t,j} - \bar{Y}_{tw})^2} \end{aligned} \quad (8)$$

where

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

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

It is important to emphasise that the asymptotic properties of the smoother employed to estimate $m(y_t)$ 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 while, in contrast, data are often characterized by spatial dependence.

Very recently, Gerolimetto and Magrini (2009) propose a two step procedure for nonparametric regression with spatially dependent data (hereafter, SNP) that does not require *a priori* parametric assumptions on spatial dependence since information on its structure is drawn from a nonparametric estimate of the errors spatial covariance matrix.

In what follows, we investigate the evolution of per capita Gross Domestic Product (GDP) inequality among European FUR through a cross-sectional distribution dynamics analysis that employs the SNP procedure with the aim of shedding some light on the consequences of neglecting spatial dependence within this framework. In the second Section we present the procedure we employ; in the third, we present the application on GDP data.

2 A new procedure for nonparametric regression with spatially dependent errors

The idea underlying the SNP procedure we adopt here is to produce an unbiased and efficient estimate of the conditional mean so to obtain a better estimate of the conditional density for the subsequent phases of the distribution dynamics analysis. The procedure consists of the following steps:

0. *Pilot fit*: estimate $m(X)$ with a local polynomial smoother, where the bandwidth is in fact a nearest neighbour smoothing parameter such that the width of each smoothing neighbourhood covers some percentage of the data. As for the degree of the polynomial, $p = 1$ is usually considered (local linear estimator). The output is $\hat{u} = Y - \hat{m}(X)$.
1. *Nonparametric covariance matrix estimation*: obtain $\hat{V}$, the estimated spatial covariance matrix of $\hat{u}$, using the spline correlogram, a continuous nonparametric positive semidefinite estimator of the covariance function developed by Bjørnstad and Falk (2001).
2. *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 Residual Spatial Autocorrelation criterion suggested by Ellner and Seifu (2002).

For brevity's reasons, we remand to the paper by Gerolimetto and Magrini (2009) for more details on the SNP procedure.

3 Empirical analysis

We now come to the study of per capita Gross Domestic Product (GDP) disparities across European regions. Due mainly to the availability of data, administratively defined regions are commonly used in empirical analyses. Within the European context, the typical example is represented by the Nomenclature of Territorial Units for Statistics (NUTS), a multi-level classification characterised by a profound heterogeneity at every level, being the result of the unification of the regional systems already existing in EU Member countries. However, one fundamental problem arises from the use of administratively defined regions in the present context. Since output is measured at workplaces while population at residences, unless the definition of a region has been selected to abstract from commuting patterns, the measured levels and growth rates of per capita GDP will be highly misleading (Cheshire and Magrini, 2009; Magrini, 2004 and 2009). Consequently, we analyse convergence in per capita GDP across a set of Functional Urban Regions (FURs), i.e. units that are broadly similar in concept to the Metropolitan Statistical Areas used in the US. In particular, FURs are defined on the basis of core cities identified by concentrations of employment and hinterlands from which more commuters flow to the employment core than to any other subject to a minimum cut off (Cheshire and Hay, 1989).

In this paper, we use data on per capita GDP at US\$ PPPs from 1977 to 2005 for 124 major Functional Urban Regions (belonging to EU12 plus Sweden). Table 1 shows the results of Moran's I test on the data where the spatial weights matrix has been constructed using a nearest-neighbours percentage equal to 10% starting from a distance matrix referring to road distances measured in minutes and inference is based on 9999 random permutations. For both years, the test finds evidence of significant spatial dependence.

Table 1 Moran's I test on per capita GDP

	1977	2005
Moran's I	0.4807	0.1707
p-level	0.0001	0.0001

Next, we estimate the mean function involved in the estimation of the conditional density using three different nonparametric techniques:

- the local constant estimator (hereafter, NP0)
- the local linear estimator (hereafter, NP1)
- the SNP procedure, with a local polynomial estimator of degree 1 (hereafter, SNP1) as described in Section 2

Figure 1 displays the estimated mean functions obtained from the three different techniques which makes the consequences of neglecting spatial dependence quite evident. Table 2 shows the outcome of the Moran's I test on regression residuals

using the same spatial weight matrix as in the previous test. The table reveals that only the SNP1 residuals show evidence against spatial dependence. In short, this means that this procedure is effectively able to incorporate the information on spatial dependence so that none is left in the residuals.

Table 2 Moran's I test on regression residuals

	NP0	NP1	SNP1
Moran's I	0.0614	0.1100	0.0197
p-level	0.0279	0.0017	0.1757

In Figure 2 we present the results of the distribution dynamics analysis. In particular, the figure is made up of three vertical sections corresponding respectively to the three techniques for the estimation of the mean function. Each section, in turn, consists of a 3-dimensional plot of the estimated conditional density (top pane), the corresponding contour plot (middle pane) and, finally, a plot of the implied stationary distribution (bottom pane). In order to interpret these figures, it should be noted that the main diagonal in the first two plot types highlights persistence properties: when most of the graph is concentrated along this diagonal, then elements in the cross-sectional distribution remain where they started. In contrast, a tendency towards convergence to equality would be signalled by a concentration of most of the graph around the 1-value of the time $t + s$ axis and parallel to the time t axis. As for the plot of the stationary distribution, when the graph displays a tendency towards a point mass, then we can conclude that there is convergence towards equality. If, on the other hand, the stationary distribution exhibits a tendency towards a two-point or bimodal measure, one can interpret this as a manifestation of income polarisation.

Section (a) of the figure, referring to a mean function estimated with NP0, conveys evidence of convergence: there is a clear clock-wise shift of the probability mass for per capita GDP values in excess of 1 in 1977 and, consistently, the stationary distribution (the dashed line) is more sharply peaked around the mean with

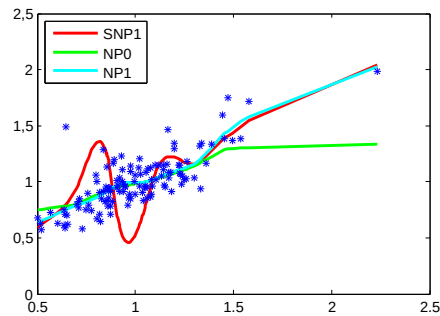


Fig. 1 Mean functions

a lighter right tail. Moving to section (b), corresponding to the use of NP1, the estimated conditional density is closely concentrated along the main diagonal, thus suggesting strong persistence. This is confirmed by a stationary distribution that overlaps almost completely the 1977 distribution. Finally, section (c) shows the implications of using SNP1 for the estimate of the mean function. More specifically, the ups and downs in the estimated conditional density indicate more complex dynamics than those revealed by the previous estimates and the two peaks characterising the stationary distribution suggest the existence of a process of polarisation among European regions.

As a conclusion, the comparison between sections (a) and (b) of Figure 2 shows that neglecting the mean bias might affect the results as the stationary distribution corresponding to NP0 emphasises convergence dynamics that are not implicit in the stationary distribution corresponding to NP1 that, instead, suggests a pattern of strong persistence. More importantly from our viewpoint, the comparison between sections (b) and (c) of the same figure shows that neglecting spatial dependence

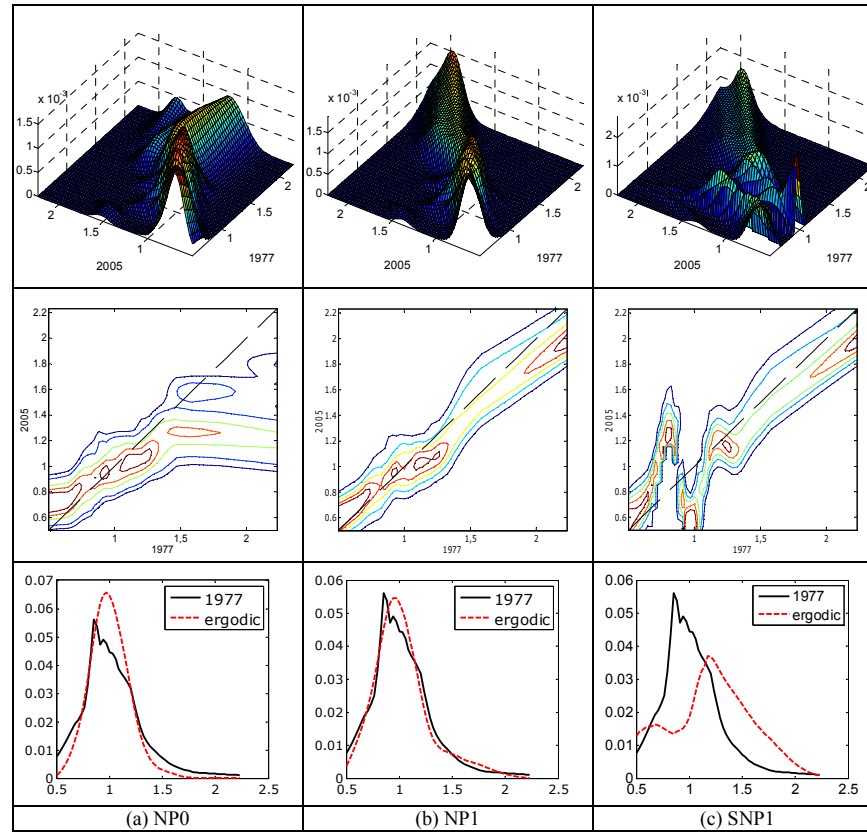


Fig. 2 Distribution dynamics

might affect the results: the result that can be obtained when SNP1 is employed in estimating the mean function is at odds with those obtained when more traditional estimators are employed. It only in section (c) that a pattern of polarisation clearly emerges. Intuitively, this could be due to the fact that the inclusion of the information on spatial dependence in the estimation process allows to recognise more clearly the relative behaviour of contiguous regions.

References

- Azzalini, A., Bowman A.W.: *Applied Smoothing Techniques for Data Analysis*. Clarendon Press, Oxford (1997)
- Bjørnstad, O.N., Falck, W.: Nonparametric Spatial Covariance Functions: Estimation and Testing. *Environmental and Ecological Statistics* **8**, 53–70 (2001)
- Cheshire, P.C., Hay, D.G.: *Urban Problems in Western Europe: an economic analysis*. Unwin Hyman, London (1989)
- Cheshire, P.C., Magrini, S.: Urban Growth Drivers in a Europe of Sticky People and Implicit Boundaries. *Journal of Economic Geography* **9**, 85–115 (2009)
- Durlauf, S.N., Johnson, P.A., Temple, J.R.W.: Growth Econometrics. In: Aghion, P., Durlauf, S.N. (eds.) *Handbook of Economic Growth*, Volume 1. Elsevier, Amsterdam (2005)
- Durlauf, S.N., Quah, D.T.: The New Empirics of Economic Growth. In: Taylor, J., Woodford, M. (eds.) *Handbook of Macroeconomics*. Elsevier, Amsterdam (1999)
- Ellner, S.P., Seifu, Y.: Using Spatial Statistics to Select Model Complexity. *Journal of Computational and Graphical Statistics* **11**, 348–369 (2002)
- Gerolimetto, M., Magrini, S.: Nonparametric Regression with spatially dependent data. DSE Working Paper 20/2009, Ca' Foscari University of Venice (2009).
- Hyndman, R.J., Bashtannyk, D.M., Grunwald, G.K.: Estimating and Visualizing Conditional Densities. *Journal of Computational and Graphical Statistics* **5**, 315–336 (1996)
- Magrini, S.: Regional (Di)Convergence. In: Henderson, J.V., Thisse J.F. (eds.) *Handbook of Regional and Urban Economics*, Volume 4. Elsevier, Amsterdam (2004)
- Magrini, S.: Why Should We Analyse Convergence Using the Distribution Dynamics Approach? *Scienze Regionali - Italian Journal of Regional Science* **8**, 5–34 (2009)
- Quah, D.T.: Empirical Cross-section Dynamics in Economic Growth. *European Economic Review* **37**, 426–434 (1993)
- Quah, D.T.: Empirics for Economic Growth and Convergence. *European Economic Review* **40**, 1353–1375 (1996)
- Quah, D.T.: Empirics for Growth and Distribution: Stratification, Polarization, and Convergence Clubs. *Journal of Economic Growth* **2**, 27–59 (1997)