Convergence analysis as distribution dynamics when data are spatially dependent

Margherita Gerolimetto
University of Venice

Stefano Magrini
University of Venice

First Draft: July 2010

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 and U.S. Metropolitan Statistical Areas through a distribution dynamics approach in which the conditional mean is estimated via a procedure that allows for spatial dependence (Gerolimetto and Magrini, 2009).

Address for correspondence:
Stefano Magrini
Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta S.Giobbe
30121 Venezia - Italy
Phone: (++39) 041 2349194
Fax: (++39) 041 2349176
e-mail: email

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 whose per capita income (relative to the group average) at time t and $t+s$ is described, respectively, by the random variables X and Y. Now, denote with $F(X)$ and $F(Y)$ the distribution of X and Y and, assuming they admit a density, indicate these with $f(X)$ and $f(Y)$. Finally, assume that the dynamics of $F(\bullet)$, or equivalently of $f(\bullet)$, can be modelled as a first order process. As a result, the density prevailing at time $t + s$ is given by

$$f\left(Y\right) = \int_{-\infty}^{\infty} f\left(Y|X\right) \ f\left(X\right) \ dX \tag{1}$$

where the stochastic kernel $f(Y|X)$ 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 $f(Y)$ as $s \to \infty$, 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.

---

[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.

Effectively, the stochastic kernel in equation (1) is a conditional density function, a nonparamentric estimate of which can be obtained by dividing the estimate of the joint probability density function $f(X, Y)$ by the estimate of the marginal probability density function $f(X)$:

$$\hat{f}(Y|X) = \frac{\hat{f}(X, Y)}{\hat{f}(X)} \tag{2}$$

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 an explanatory and a dependent variable, respectively; also, denote by $\{(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)\}$ a sample of length n, and by $\{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$ 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 density of Y conditional on $X = x$ is:

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

where

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

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

---

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

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*.

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

By construction, the mean-bias of the estimator in (7) is equal to the bias of 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 smoother employed to estimate $m(x)$ are based on the assumption that the error terms of the nonparametric

4

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 non-parametric 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 per capita GDP regional disparities. In particular, we use data on per capita GDP at US$ PPPs from 1995 to 2005 for 124 major Functional Urban Regions (belonging to EU12 plus Sweden) and data on per capita GDP from 2001 to 2008 for 363 Metropolitan Statistical Areas belonging to the contiguous United States. To do this we conduct for both data sets 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 of the paper we present the procedure we employ; in the third, we present the application on per capita 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.

Along the lines of Martins-Filho and Yao (2009), the procedure moves from a pilot estimate of the nonparametric regression with the local linear estimator and consists of two steps: firstly, the spatial covariance matrix is estimated nonparametrically through a spline correlogram (Bjørnstad and Falk, 2001); secondly, a modified regression is run exploiting the information on spatial dependence just obtained.

The procedure consists of the following steps [3]:

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 spatial 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).

The idea developed in this paper is that the SNP procedure can be employed to estimate the mean function representing the correct input to be supplied to the distribution dynamics analysis. More precisely, in case there is the suspicion that data can be affected by spatial dependence, the mean bias adjustment introduced by Hyndman to estimate conditional densities should be based on an estimate of the mean function obtained using SNP. Hence, the resulting distribution dynamics' estimates are free from the effects of the spatial dependence.

---

[3]Refer to Gerolimetto and Magrini (2009) for more details on the SNP procedure.

3 Empirical analysis

We now come to the study of per capita GDP regional disparities. Due mainly to the availability of data, administratively defined regions are commonly used in empirical analyses. However, as emphasised elsewhere (Cheshire and Magrini, 2009; Magrini, 2004 and 2009), administratively defined regions are likely to misrepresent per capita income dynamics of the underlying economies. On the one hand, 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 can be highly misleading. On the other hand, resorting to relatively large regions, such as the States in the U.S. context or the NUTS-1[4] regions in Europe, does not represent a solution as it simply tends to 'aggregate away' truly spatial differences.

Consequently, we analyse regional convergence in per capita GDP using two sets of functionally defined regions, Functional Urban Regions (FURs) for Europe and Metropolitan Statistical Areas (MSAs) for the United States. The two sets are broadly similar in concept as their basic units contain a large urban core and a hinterland characterised by a high degree of social and economic integration (as measured by commuting flows) with the urban core[5]. In particular, in this paper we use data on per capita GDP at U.S.\$ PPPs from 1995 to 2005 for 124 major Functional Urban Regions (belonging to EU12 plus Sweden) and data on per capita GDP from 2001 to 2008 for 363 Metropolitan Statistical Areas belonging to the contiguous U.S. states.

We first analyse the convergence pattern among FURs. Table 1 shows the results of Moran's I test on FURs' data where the spatial weights matrix has been constructed using 10 nearest-neighbours starting from a distance matrix referring to road distances measured in minutes and inference is based on 9999 random permutations. For both

---

[4]NUTS stands for Nomenclature of Territorial Units for Statistics

[5]The main difference is that in the definition of FURs more emphasis is placed on the concentration of jobs. For details see Cheshire and Hay (1989).

years, the test finds evidence of significant spatial dependence.

Table 1: Moran's I test on per capita GDP (FURs)

	1995	**2005**
Moran's I	0.3849	0.1925
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), where the bandwidth parameter is set to 0.40

- the local linear estimator (hereafter, NP1), where the bandwidth parameter is set to 0.40

- the SNP procedure, with a local polynomial estimator of degree 1 (hereafter, SNP1) as described in Section 2; the bandwidth parameter for the pilot fit is equal to 0.03 and the final bandwidth is 0.45^6.

Figure 1 displays the estimated mean functions obtained from the three different techniques and makes the consequences of neglecting spatial dependence quite evident. Moreover, the estimate obtained using NP0 clearly suffers from boundary bias, which is removed when using both NP1 and SNP1. 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.

In Figures 2, 3 and 4 we present the results of the distribution dynamics analysis. In particular, each figure is made up of three graphs representing the 3-dimensional plot of

---

[6]Udersmoothing in the pilot estimate is common practice in two-step procedures in order to avoid piling-up of the bias.

Table 2: Moran's I test on regression residuals (FURs)

	NP0	NP1	SNP1
Moran's I	0.0631	0.2605	0.0247
p-level	0.0308	0.0001	0.1697

the estimated conditional density (top left), the corresponding contour plot (top right) and, finally, a plot of the implied stationary distribution (bottom). In order to interpret these figures, it should be noted that the main diagonal in the two upper plots highlights persistence properties: when most of the graph is concentrated along this diagonal, then elements in the cross-sectional distribution tend to 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.

Figure 2 shows distribution dynamics estimated using the traditional kernel density estimator (NP0)[7]. The shape of the conditional density suggests a tendency to converge for most regions; however, regions with a level of relative per capita GDP in excess of 1.25 tend to split into two groups, one converges to the rest of the distribution and the other converges to a separate cluster. The effect of this behaviour is clearly evident from the plot of the stationary distribution that features a small club of rich regions.

In the case of a kernel density estimator with mean bias adjustment through a local linear estimator (NP1), the tendency towards convergence becomes weaker (Figure 3). In particular, there is an evident counter-clockwise rotation of the tails of the conditional probability mass with respect to the NP0 case. Consistently, the stationary distribution

---

[7]All kernel density estimates use bandwidths based on the normal reference rule. In particular, we obtain 0.0685 for 1995 and 0.0382 for 2005.

is less peaked and with slightly heavier tails. Moreover, the small club of rich regions, while still present, becomes less separated from the rest of the distribution.

Finally, what emerges from the analysis of figures corresponding to a mean bias adjustment through a spatial local linear estimator (SNP1) is a tendency towards the development of three convergence clubs (Figure 4). While the dynamics characterising richer regions are something in between what observed in the two previous representations, it is now evident that estimating the mean function allowing for spatial dependence makes visible the presence of a third club populated by poor regions. In addition to the main central club and a small club of rich regions, the stationary distribution thus shows a sizeable club of poor regions.

Now we come to the analysis of distribution dynamics for the Metropolitan Statistical Areas between 2001 and 2008. As before, we begin with the analysis of spatial dependence. Table 3 then shows Moran's I test results for MSAs where the spatial weights matrix has been constructed using 10 nearest-neighbours starting from a matrix of euclidean distances across regional centroids and inference is based on 9999 permutations. Also in this case, the test finds evidence of significant spatial dependence in both years.

Table 3: Moran's I test on per capita GDP (MAs)

	2001	2008
Moran's I	0.0520	0.0452
p-level	0.0111	0.0197

Figure 5 displays the estimated mean functions obtained from the three different techniques using the following values for the estimation parameters: for NP0 and NP1 the bandwidth parameter is set to 0.40; for the SNP procedure the bandwidth parameter is equal to 0.02 in the pilot fit and to 0.45 the final fit. Also in this case, the consequences of neglecting spatial dependence are quite evident as well as the effects of the boundary bias suffered by NP0. In addition, Table 4 shows the outcome of the Moran's I test on re-

gression 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 also in this case the SNP1 procedure is effectively able to incorporate the information on spatial dependence so that none is left in the residuals.

Table 4: Moran's I test on regression residuals (MSAs)

	NP0	NP1	SNP1
Moran's I	0.0698	0.1389	0.0233
p-level	0.0034	0.0001	0.1245

Figure 6, referring to a mean function estimated with NP0[8], conveys evidence of weak convergence. The stationary distribution appears to have lighter tails than the initial one while the two peaks already present in the main part of the distribution eventually become closer. There is also a small bump on the right tail, presumably corresponding to the behaviour of Bridgeport-Stamford-Norwalk (CT) and San Jose-Sunnyvale-Santa Clara (CA), whose emergence is explained by the dynamics appearing in the top-right part of the contour plot.

Moving to Figure 7 that corresponds to the use of NP1, the contour lines of the estimated conditional density are signalling that the probability mass is closely concentrated along the main diagonal implying that the evolution of the cross-sectional distribution is quite a slow process. Another striking feature of the estimated conditional density is the presence of two peaks at the two extremes of the 3-dimensional representation that act as attractors. The consequences of this are very evident in the stationary distribution which shows two clearly distanced clubs at the two extremes of the support.

Finally, Figure 8 shows the implications of using SNP1 for the estimate of the mean function. More specifically, the two clubs detected in the previous case are now less distanced and, possibly, more connected by a third, centrally located, club. In addition,

---

[8]As before, bandwidth in the kernel estimates are based on the normal reference rule: 0.0454 for the 2001 series and 0.0274 for 2008.

the lowest income club is quite sharply peaked, confirming a behaviour which is also evident in the conditional density graphs.

4 Conclusions

In this paper we have studied the evolution of per capita GDP inequalities among European Functional Urban Regions between 1995 and 2005 and among U.S. Metropolitan Statistical Areas between 2001 and 2008. Specifically, we have employed a distribution dynamics approach in which the conditional mean is estimated via a procedure that allows for spatial dependence and compared its results with those obtained from more conventional estimators.

In general, the comparison between estimated conditional densities and stationary distributions confirms that neglecting spatial dependence might substaintially affect the results. Moreover, the comparison of the different mean functions – as well as of the corresponding conditional densities – clearly suggests that the boundary bias characterising the local constant estimator (NP0) translates into a bias towards convergence in the distribution dynamics analysis.

Overall, we detect evidence of club convergence for both sets of regions, appearing to be particularly strong among Metropolitan Statistical Areas.

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. Enviromental 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)

Figures
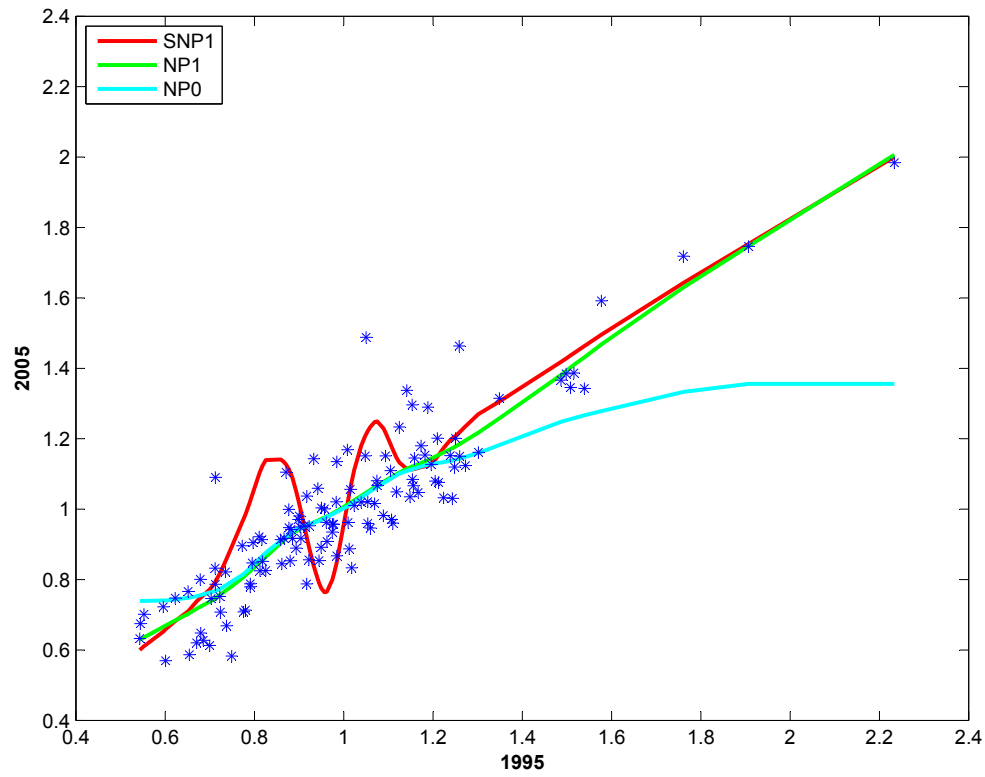
Figure 1: Mean functions (FURs)
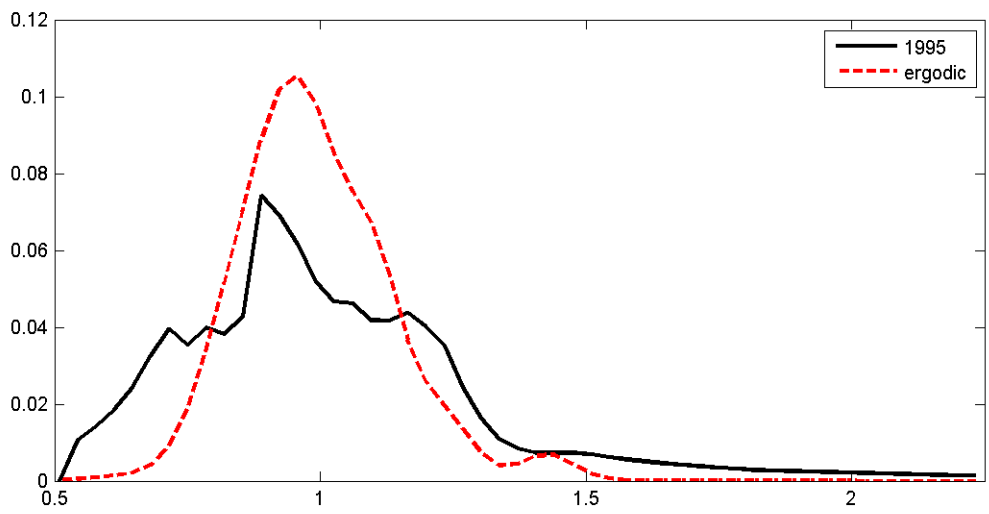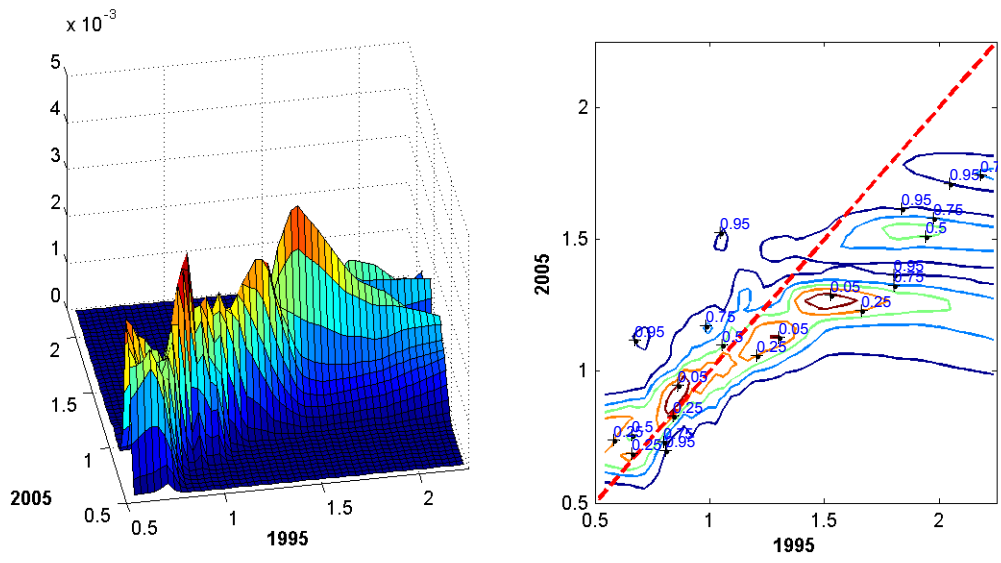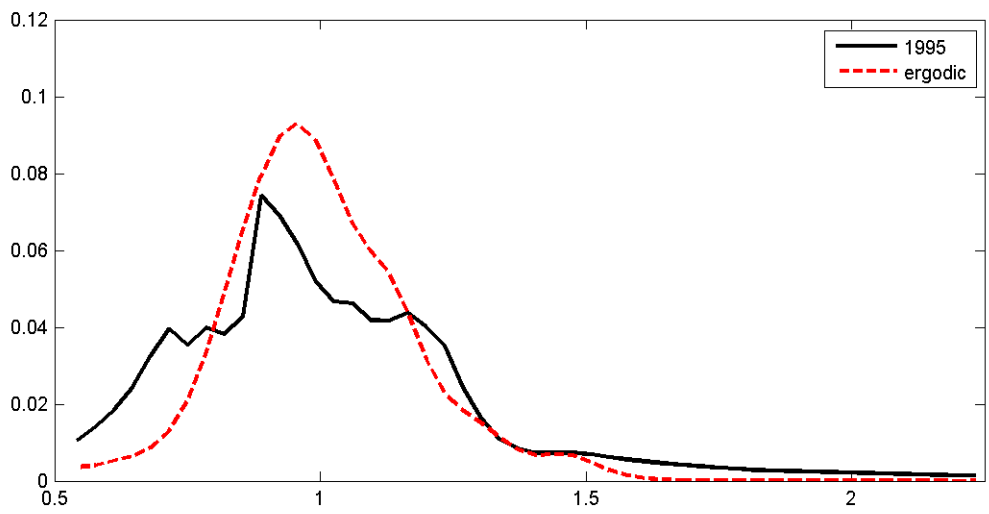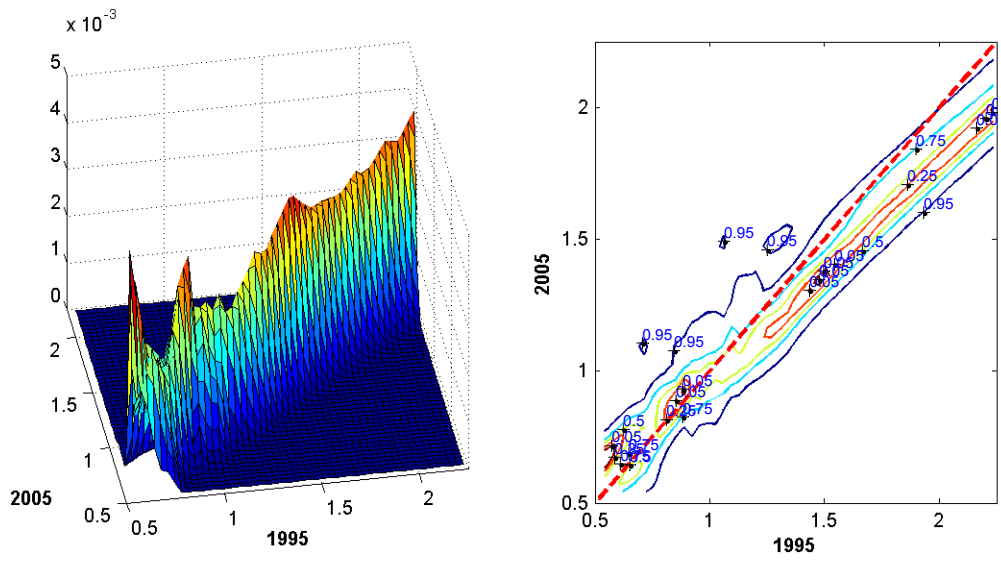
Figure 2: Distribution dynamics (FURs - NP0)
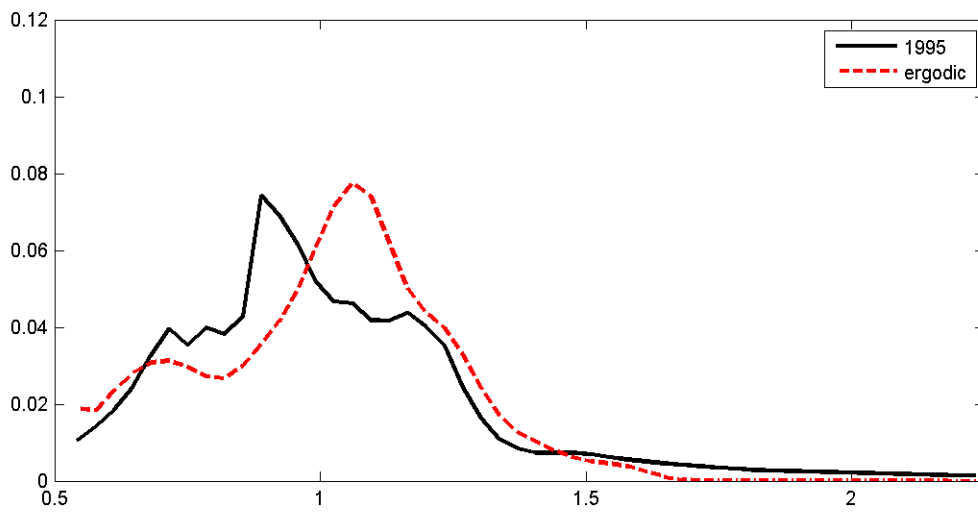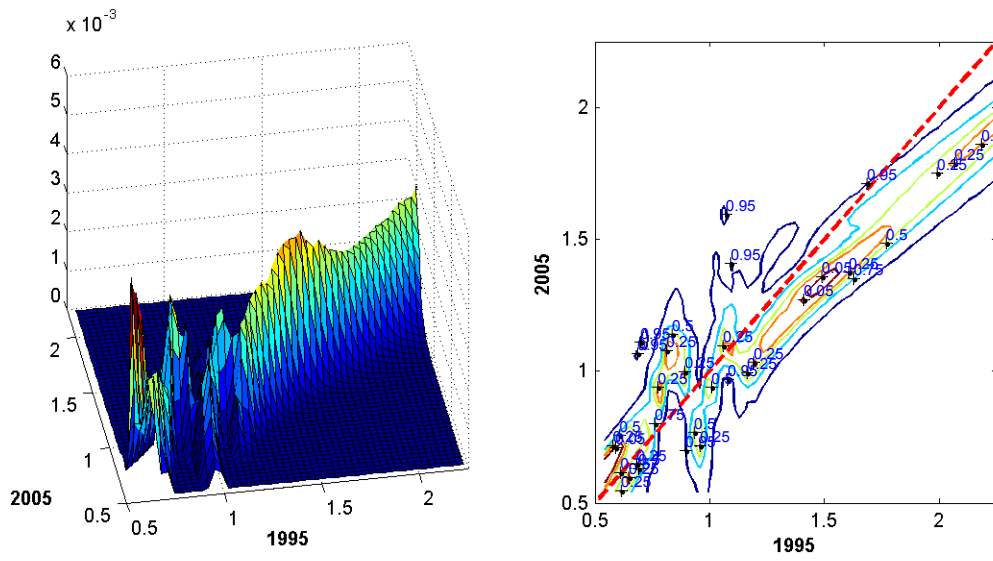
Figure 3: Distribution dynamics (FURs - NP1)
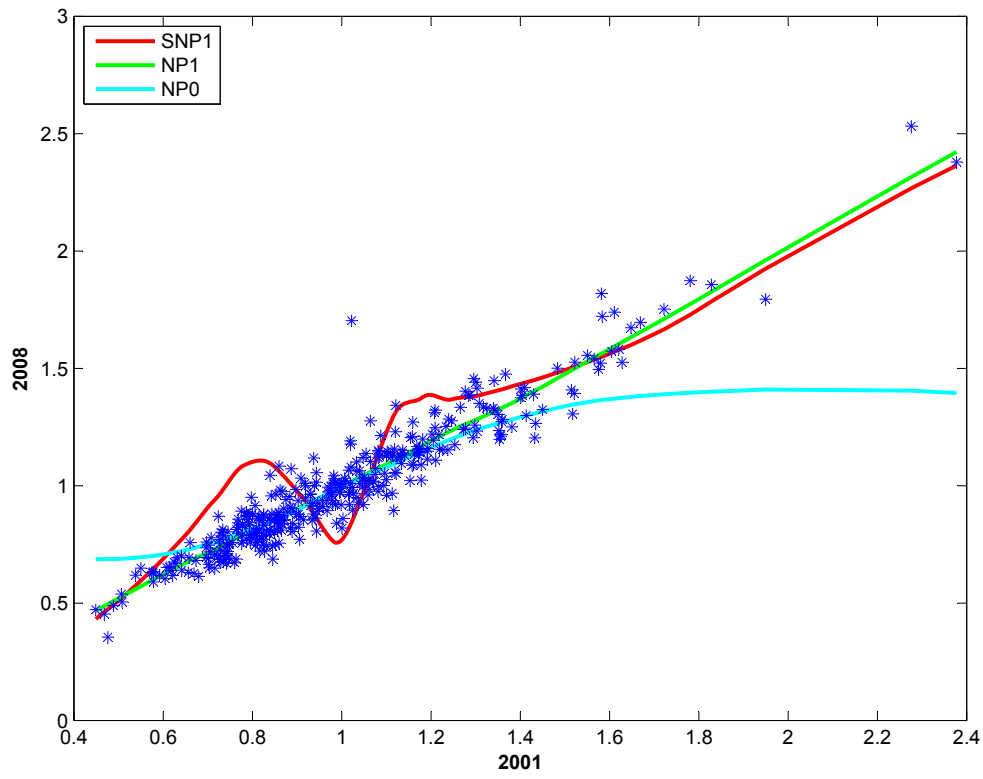
Figure 4: Distribution dynamics (FURs - SNP1)
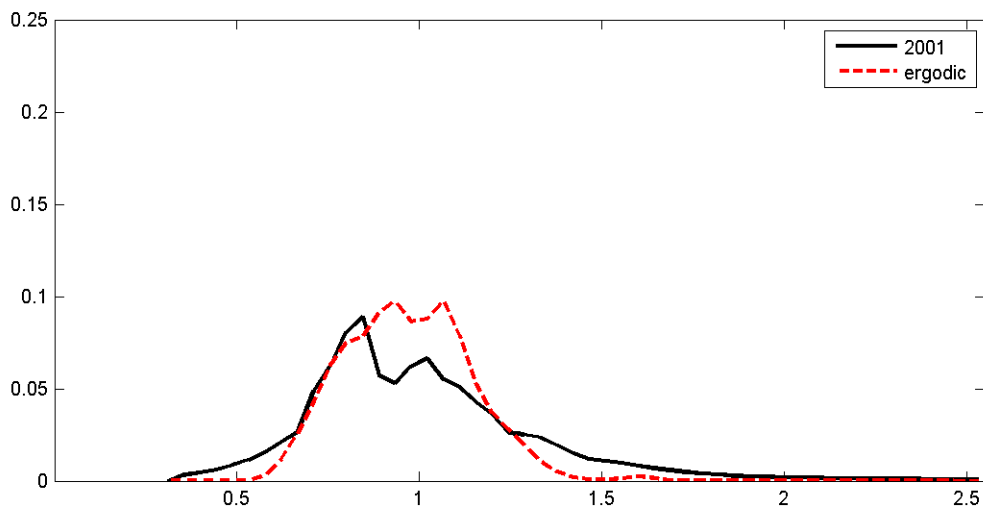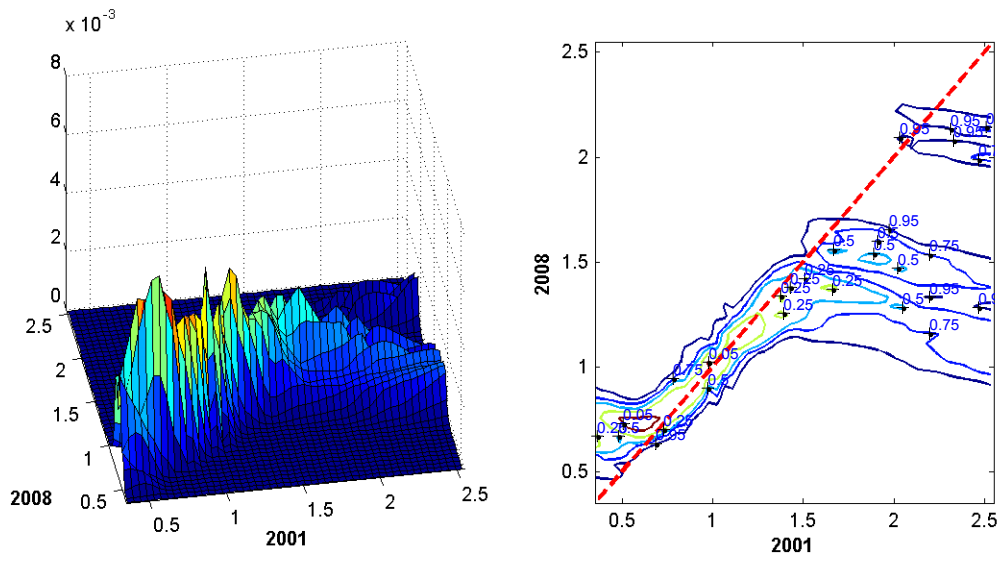
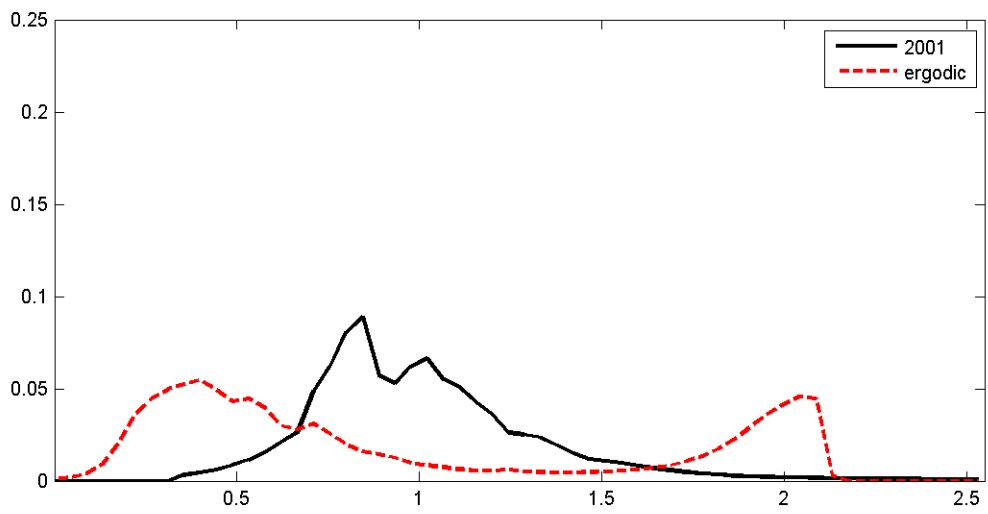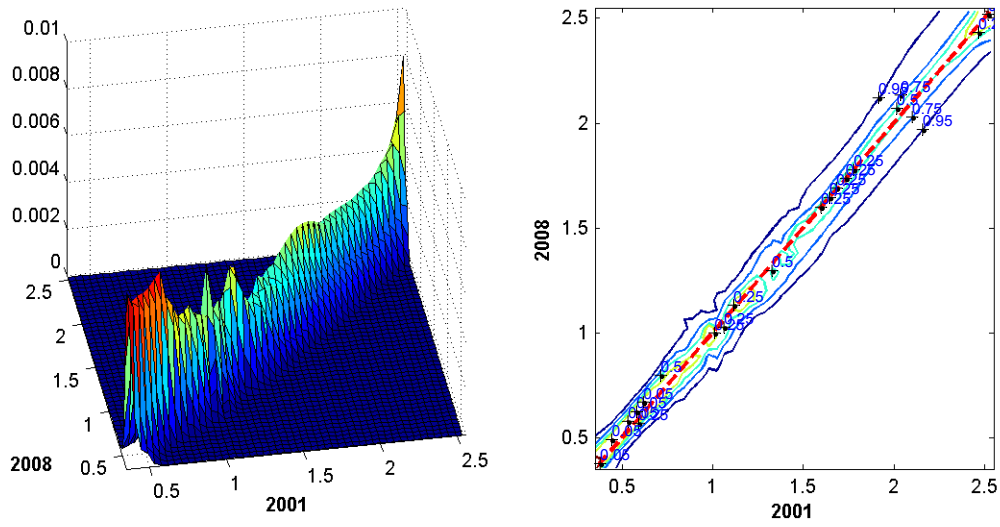Figure 5: Mean functions (MSAs)

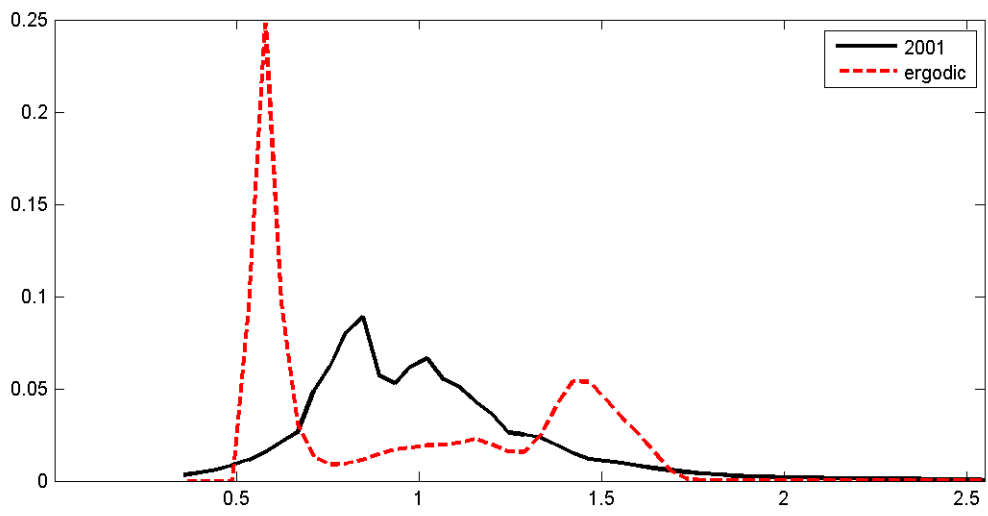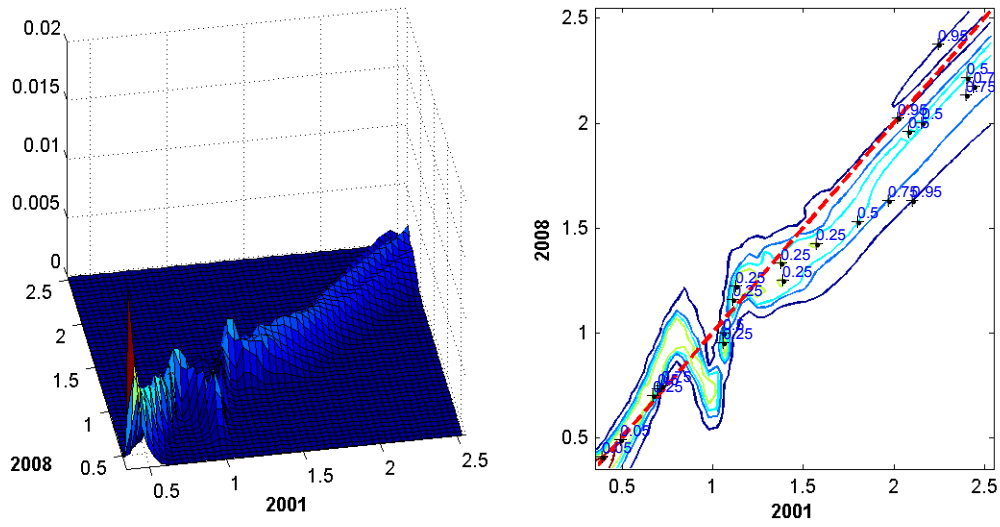Figure 6: Distribution dynamics (MSAs - NP0)

Figure 7: Distribution dynamics (MSAs - NP1)

Figure 8: Distribution dynamics (MSAs - SNP1)