



Ca' Foscari
University
of Venice

Department
of Economics

Working Paper

Margherita Gerolimetto
and
Stefano Magrini

Distribution Dynamics in
the US.
A Spatial Perspective

ISSN: 1827-3580
No. 02/WP/2016





Distribution Dynamics in the US. A Spatial Perspective

Margherita Gerolimetto
Ca' Foscari University Venice

Stefano Magrini
Ca' Foscari University Venice

22 January 2016

Abstract

It is quite common in cross-sectional convergence analyses that data exhibit strong spatial dependence. While the literature adopting the regression approach is now fully aware that neglecting this feature may lead to inaccurate results and has therefore suggested a number of statistical tools for addressing the issue, research is only at a very initial stage within the distribution dynamics approach. In particular, in the continuous state-space framework, a few authors opted for spatial pre-filtering the data in order to guarantee the statistical properties of the estimates. In this paper, we follow an alternative route that starts from the idea that spatial dependence is not just noise but can be a substantive element of the data generating process. In particular, we develop a tool that, building on a mean-bias adjustment procedure established in the literature, explicitly allows for spatial dependence in distribution dynamics analysis thus eliminating the need for pre-filtering. Using this tool, we then reconsider the evidence on convergence across US states.

Keywords

Distribution Dynamics, Nonparametric Smoothing, Spatial Dependence.

JEL Codes: C14, C21.

Address for correspondence:

Margherita Gerolimetto
Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta S.Giobbe
30121 Venezia - Italy
Phone: (+39) 041 2347432
Fax: (+39) 041 2349210
e-mail: margherita.gerolimetto@unive.it

This Working Paper is published under the auspices of the Department of Economics of the Ca' Foscari University of Venice. Opinions expressed herein are those of the authors and not those of the Department. The Working Paper series is designed to divulge preliminary or incomplete work, circulated to favour discussion and comments. Citation of this paper should consider its provisional character

Distribution Dynamics in the US. A Spatial Perspective

Margherita Gerolimetto and Stefano Magrini

Abstract

It is quite common in cross-sectional convergence analyses that data exhibit strong spatial dependence. While the literature adopting the regression approach is now fully aware that neglecting this feature may lead to inaccurate results and has therefore suggested a number of statistical tools for addressing the issue, research is only at a very initial stage within the distribution dynamics approach. In particular, in the continuous state-space framework, a few authors opted for spatial pre-filtering the data in order to guarantee the statistical properties of the estimates. In this paper we follow an alternative route that starts from the idea that spatial dependence is not just noise but can be a substantive element of the data generating process. In particular, we develop a tool that, building on a mean-bias adjustment procedure established in the literature, explicitly allows for spatial dependence in distribution dynamics analysis thus eliminating the need for pre-filtering. Using this tool, we then reconsider the evidence on convergence across US states.

Keywords: Distribution dynamics, Nonparametric smoothing, Spatial dependence

JEL Codes: C14, C21

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 by Paelink and Klaassen (1979), Bartels and Ketelapper (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 regional per capita income convergence, as in the present paper. However, while in the literature on convergence through the regression approach there is full awareness that neglecting spatial dependence may lead to biased and inefficient estimates (e.g. Rey and Montouri, 1999), this issue has so far received much less attention within the literature that adopts, as we do here, the distribution dynamics approach. Typically, within the latter approach the issue is tackled by adopting a spatial filtering technique 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 G_i developed by Getis and Ord (1992). A strict assumption however underlies this way of approaching the issue: spatial dependence is seen as a nuisance element that should be eliminated in order to avoid the risk of losing the statistical properties of the estimates (Anselin, 2002 and 2009).

Differently from this view, we think that spatial dependence is often likely to be a substantive element of the process under study and this, in particular, should be the case when studying economic convergence across regional units. Just to give an example, not only it is well known that the level of per capita income in a US state is correlated to the level observed in neighboring states but, as shown by Rey (2001), also the mobility of the states within the cross-sectional distribution of per capita income is significantly affected by the relative position of geographical neighbors within the same distribution. In such instances,

spatial dependence appears to embody valuable information on convergence dynamics and adopting a spatial filtering technique represents a controversial strategy (Magrini, 2004) as it may yield misleading results.

With this concerns in mind, in this paper we propose a technique that explicitly allows for spatial dependence in distribution dynamics analysis thus eliminating the need for pre-filtering. In simple terms, the distribution dynamics approach analyses the evolution of a variable's distribution by means of a stochastic kernel, effectively a conditional density function, which is commonly estimated using the kernel density estimator. However, Hyndman *et al.*, (1996) suggest that this estimator might have poor bias properties and develop an adjustment procedure based on the estimate of a conditional mean function characterized by better bias properties. Here, we exploit this idea further and enrich the estimate of the conditional density through 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 nonparametric regression estimator that moves from the standard local linear estimator. More in details, we draw on the work by Martins-Filho and Yao (2009) who establish a set of sufficient conditions for the asymptotic normality of the local linear estimator and propose a two step procedure for nonparametric regression with spatially dependent data that does not require *a priori* parametric assumptions on spatial dependence. Information on its structure is in fact drawn from a nonparametric estimate of the errors spatial covariance matrix. The finite sample performance of this estimator, called Spatial Nonparametric estimator, is then assessed via an extensive Monte Carlo experiment.

Using the Spatial Nonparametric estimator to adjust the mean-bias (Hyndman *et al.*, 1996), we are able to analyze convergence with the distribution dynamics without prefiltering. Through this novel version of distribution dynamics approach, we then reconsider the evidence on convergence across regional economies in the US and also shed some light on the consequences of neglecting spatial dependence. In particular, we employ a dataset on real per capita personal income net of current transfer receipts for 48 conterminous states

containing quarterly data from the 1st quarter of 1969 to the 3rd quarter of 2012.

The paper is structured as follows. In the second Section we recall the distribution dynamics approach. In the third Section we introduce our Spatial Nonparametric estimator. The fourth Section presents the application on per capita personal income data.

2 Distribution dynamics

Distribution dynamics (Quah, 1993 a and b, 1996 a and b, 1997) represents a relatively 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 $Y_{i,t}$ per capita income of economy i at time t (relative to the group average). Next, we 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

$$f(Y_{t+s}) = \int_{-\infty}^{\infty} f(Y_{t+s}|Y_t) f(Y_t) dY_t \quad (1)$$

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 $t + s$.

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 $f(Y_t, Y_{t+s})$ by the estimate of the marginal probability density

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.

function $f(Y_t)$:

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

Originally, the most commonly adopted method to obtain such an estimate was the kernel density estimator. However, Hyndman *et al.* (1996) suggest that this popular estimator might have poor bias properties. To clarify this, denote by $\{(Y_{1,t}, Y_{1,t+s}), \dots, (Y_{j,t}, Y_{j,t+s}), \dots, (Y_{n,t}, Y_{n,t+s})\}$ a sample of length n . Indicate the conditional mean with $m(Y_t) = E(Y_{t+s}|Y_t)$ so that:

$$Y_{j,t+s}|Y_{j,t} = M(Y_{j,t}) + \epsilon_j \quad j = 1, \dots, n \quad (3)$$

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

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

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

where

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

a and b are bandwidth parameters controlling the smoothness in, respectively, the Y_t dimension and the Y_{t+s} dimension, $K_b(u) = b^{-1}K(\frac{u}{b})$ is a scaled kernel function and $K(\cdot)$ is assumed to be a real value, integrable and non negative even function.² Note that the mean of the conditional density estimator in (4) is in fact an estimator of the conditional mean function $M(Y_t)$:

$$\hat{M}(Y_t) = \int Y_{t+s} \hat{f}(Y_{t+s}|Y_t) dY_{t+s} = \sum_{j=1}^n w_j(Y_t) Y_{j,t+s} \quad (6)$$

In addition, 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 Y_t space and also in the interior, especially when the

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

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:

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

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

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 Nadaraya-Watson smoother, the estimator reverts to the traditional kernel density estimator in (4). More importantly, it also suggests that a lower mean-bias can 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).

It is important to emphasize that the asymptotic properties of the smoother employed to estimate the mean function $M(Y_t) = E(Y_{t+s}|Y_t)$ are based on the assumption that the error terms in (3) are zero mean and uncorrelated. However, as we anticipated in the introduction, it is highly unlikely that data in empirical analyses of cross-sectional convergence comply with this hypothesis as, in contrast, they normally feature spatial dependence.

To tackle this issue without resorting to prefiltering, we hence develop a two step procedure for nonparametric regression with spatially dependent data 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. This procedure is described in the next session.

3 A spatial nonparametric regression estimator

3.1 Modeling spatial dependence

As clarified in the previous section, since we are analyzing economic convergence across spatial units, the mean function estimate required in the adjustment procedure by Hyndman *et al.* (1996) is in fact an autoregression of a variable characterized by spatial dependence of a substantive, rather than nuisance, nature. In this context, we can therefore model the data generating process (DGP) according to a general parametrization known as mixed regressive-spatial regressive model (Florax and Folmer, 1992). Before introducing its formalization, we remark that for the sake of generality we switch to the conventional notation in regression framework and indicate with Y the dependent variable (in the convergence context, Y_{t+s}) and with X the independent one (again, in the convergence context, Y_t). The mixed regressive-spatial regressive model we start from therefore is

$$Y = \rho W_y Y + X\beta + W_x X\lambda + \epsilon \quad (8)$$

where $Y = \{Y_1, Y_2, \dots, Y_n\}$ and $X = \{X_1, X_2, \dots, X_n\}$ are $n \times 1$ vectors,³ $\epsilon \sim N(0, \sigma^2 I_n)$ is a $n \times 1$ vector of innovations, $-1 < \rho < 1$, $-1 < \lambda < 1$, β is a parameter and W_y and W_x are $n \times n$ spatial weights matrices whose w_{ij} elements are non negative when $i \neq j$ and zero otherwise.

Now rewrite (8) as follows

$$Y = (I - \rho W_y)^{-1}(X\beta + W_x X\lambda) + (I - \rho W_y)^{-1}\epsilon \quad (9)$$

and hence

$$\begin{aligned} Y &= M(X) + u \\ u &= \rho W_y u + \epsilon \end{aligned} \quad (10)$$

where $M(X) = (I - \rho W_y)^{-1}(X\beta + W_x X\lambda)$. Three remarks are now in order. Firstly, the error term in (9) is non-spherical due to spatial dependence. Secondly, formalizations (9)

³ The analysis that follows can be easily generalized to the case in which X is a $n \times p$ matrix and β a $p \times 1$ vector, with $p \geq 2$.

and (10) clarify that spatial dependence is a substantive element as it enters directly in the function $M(\cdot)$. Thirdly, although model (8) we started from is linear, reformulations (9) and (10) are nonlinear in their parameters. This means that, unless one knows the entire spatial dependence structure (W_y and W_x) and the multiplicative coefficients λ and ρ , nonparametric methods should be preferably employed.

An even stronger indication favoring the use of nonparametric methods arises when the assumption of linearity in model (8) is relaxed, in which case the mixed regressive-spatial regressive becomes:

$$Y = \rho W_y Y + m(X) + g(W_x X) + \epsilon \quad (11)$$

thus yielding

$$Y = (I - \rho W_y)^{-1}(m(X) + g(W_x X)) + (I - \rho W_y)^{-1}\epsilon \quad (12)$$

where the function to be estimated becomes $M(X) = (I - \rho W_y)^{-1}(m(X) + g(W_x X))$ thus coming back again to equation (10). Reformulating the mixed regressive-spatial regressive as in (10) shows that the DGP is, in fact, a very general function $M(X)$ in which on top of the nonlinearity in its parameters noted above, there is also a nonlinear relationship in the functional form that links X to Y . For these reasons, we introduce a nonparametric estimator that is designed to generalize existing nonparametric regression estimators by allowing for spatial dependence.

3.2 Nonparametric regression with dependent errors

Nonparametric regression has now become quite a standard statistical tool when the functional form is possibly of an unknown type. Indeed, given a generic model such as

$$Y = M(X) + \epsilon$$

where ϵ is the i.i.d. error term and $M(X)$ is a smooth function, linearity of $M(X)$ can not be always safely assumed. Under these circumstances, the parametric literature typically offers non linear least squares estimates, but these require to conjecture a specific functional form

with respect to which the minimization problem has to be solved. When making assumptions on the functional form of $M(\cdot)$ is not possible or not recommended, nonparametric methods represent a valuable solution.

In general, the estimate of a nonparametric regression can be obtained by means of some smoothing methods. One of the most commonly adopted estimation technique is the Nadaraya-Watson estimator (Nadaraya, 1964; Watson, 1964):

$$\hat{M}(X) = \frac{\sum_{j=1}^n K\left(\frac{X-X_j}{h}\right)Y_j}{\sum_{j=1}^n K\left(\frac{X-X_j}{h}\right)} \quad (13)$$

where h is the bandwidth, the parameter that controls the degree of smoothness. The Nadaraya-Watson estimator is in fact a special case of local polynomial regression that applies when the degree of the smoothing polynomial is 0 and for this reason the Nadaraya-Watson is also know as the Local Constant Estimator (LCE). When the degree of the smoothing polynomial is 1 instead of 0, the smoother becomes the local linear estimator (LLE):

$$\hat{M}(X) = \frac{\sum_{j=1}^n K\left(\frac{X-X_j}{h}\right)Y_j}{\sum_{j=1}^n K\left(\frac{X-X_j}{h}\right)} + (X - \bar{X}_w) \frac{\sum_{j=1}^n K\left(\frac{X-X_j}{h}\right)(X_j - \bar{X}_w)Y_j}{\sum_{j=1}^n K\left(\frac{X-X_j}{h}\right)(X_j - \bar{X}_w)^2} \quad (14)$$

where

$$\bar{X}_w = \frac{\sum_{j=1}^n K\left(\frac{X-X_j}{h}\right)X_j}{\sum_{j=1}^n K\left(\frac{X-X_j}{h}\right)}$$

Similarly to the parametric regression environment, nonparametric regression estimators generally assume i.i.d. error terms. In case of lack of independence, Robinson (2008, 2011) derives consistency and asymptotic distribution theory for the local constant regression estimator in relation to various kinds of spatial data. 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 and heteroschedastic. 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 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, attention within the nonparametric literature has focussed on estimators that, by incorporating the information contained in the error covariance structure, outperform, both asymptotically and in finite samples, traditional nonparametric ones. In particular, Martins-Filho and Yao (2009) develop a two-step procedure whose asymptotic validity is proved under rather general covariance structures.

More formally, Martins-Filho and Yao consider the following nonparametric regression:

$$Y = M(X) + u \tag{15}$$

where the error term u is such that $E(u_i) = 0$, $\forall i = 1, \dots, n$, and $E(u_i, u_j) = \omega_{ij}(\theta_0)$, $\theta_0 \in R^p$, $p < \infty$ and demonstrate (Martins-Filho and Yao, 2009; Theorem 2) the asymptotic normality and convergence rate of the traditional LLE of model (15). In addition, they observe that this estimator, \hat{M} , does not exploit the information contained in the error term correlation structure. Therefore, to improve its performance, they suggest a two-step procedure that incorporates this information in order to yield spherical error terms. More in detail, let $\Omega(\theta_0)$ denote the $n \times n$ matrix with elements ω_{ij} and $P(\theta_0)$ be a $n \times n$ matrix such that $\Omega(\theta_0) = P(\theta_0)P(\theta_0)'$. Now, by defining the new regressand as $Z = P(\theta_0)^{-1}Y + (I_n - P(\theta_0)^{-1})M(X)$, Martins-Filho and Yao replace the original regression with the following

$$Z = M(X) + \epsilon \tag{16}$$

where the error terms $\epsilon = P(\theta_0)^{-1}u$ are now spherical by construction. The new estimator, $\check{M}(X)$, is simply the LLE of (16). With an additional assumption constraining the nature of the stochastic process \mathbf{u} to be a linear transformation of i.i.d. processes, the authors show $\check{M}(X)$ to represent an improvement over $\hat{M}(\mathbf{X})$ in terms of efficiency (Martins-Filho and Yao, 2009; Theorem 3). To guarantee the bias from the first stage estimator to be smaller

than the leading bias coming from the second stage, as is usual in the literature on two-stage nonparametric regression, undersmoothing in the first stage is required.

Since Z is not observed (it depends on the unknown $M(X)$ and $P(\theta_0)$), Martins-Filho and Yao propose a feasible version of the $\check{M}(\mathbf{X})$ estimator. This estimator, $\dot{M}(X)$, is based on an observed regressand

$$\dot{Z} = P(\hat{\theta})^{-1}Y + \left(I_n - P(\hat{\theta}^{-1})\right) \hat{M}(X)$$

where a pilot local linear estimate $\hat{M}(X)$ is used in place of $M(X)$ and $P(\hat{\theta})$ in place of $P(\theta_0)$. The authors also provide an asymptotic result⁴ that guarantees that, as long as a consistent estimate $\hat{\theta}$ is plugged into $P(\theta_0)$, the feasible estimator $\dot{M}(X)$ is asymptotically equivalent to $\check{M}(X)$.

3.3 A new nonparametric regression estimator for spatially dependent data

Building on the theoretical background of the two-step nonparametric regression estimator by Martins-Filho and Yao, we can now introduce our proposal of a spatial nonparametric regression estimator (SNP).

Since spatial dependence has not been included among the forms of dependence considered by Martins-Filho and Yao, here further assumptions need to be made on the form of the spatial covariance matrix of the error term. In particular, apart from adopting assumptions A1-A6 characterising the original framework developed by these authors (Martins-Filho and Yao, 2009; page 311 and 313), we: *i.* suppose the error term to possess a spatial covariance matrix such that the weighted average of the main diagonal elements converge as $n \rightarrow \infty$ and *ii.* impose spatial mixing conditions, as in Jenish and Prucha (2009).⁵

A peculiar feature of our procedure is that the spatial covariance matrix is estimated nonparametrically starting from a direct representation of spatial dependence. The logic

⁴ Martins-Filho and Yao (2009), Theorem 4

⁵ Note that Cliff-Ord models (Cliff and Ord, 1973) trivially meet these assumptions.

underlying the estimate of the covariance matrix of the error term is that, unless the form of dependence is of interest itself, it is better not to parametrize it. From this viewpoint, Robinson (1987) estimates the residuals variance, conceived as an unknown function of the explanatory variables, by a nearest neighbors nonparametric regression and proves that asymptotic properties of the estimated residuals variance (as well as that of other parameters involved in the multiple regression) are guaranteed if k the number of nearest neighbors, increases slowly with the sample size. In developing our estimator, we follow Robinson's logic relatively to a multivariate regression with non spherical errors and estimate consistently the spatial covariance matrix through a nonparametric methodology, called spline correlogram, whose details are presented in the next subsection.

3.3.1 Nonparametric estimation of the spatial covariance matrix

A commonly adopted approach to express the elements of a generic spatial covariance matrix Ω is through a direct representation of the dependence as some function of the distance separating sites s_i and s_j . In such an instance, the spatial autocovariance function is defined by

$$\gamma(s_i, s_j) = \sigma^2 f(d_{ij}, \phi) \quad (17)$$

and the spatial autocorrelation function by

$$\rho(s_i, s_j) = f(d_{ij}, \phi) \quad (18)$$

where d_{ij} is the distance between sites i, j and $f(\cdot)$ is a decaying function such that $\frac{\partial f}{\partial d_{ij}} < 0$, $|f(d_{ij}, \phi)| \leq 1$ with ϕ being an appropriate vector of parameters. Within this framework, the spatial covariance matrix Ω is positive definite and composed by elements ω_{ij} , obtained through function $\gamma(s_i, s_j)$ in observed distances across sites. These features of matrix Ω follow directly from the stationarity and isotropy assumptions underpinning the existence of a spatial covariance function.⁶

⁶ These assumptions are certainly met when Ω represents the spatial covariance matrix of homoskedastic errors of a Cliff-Ord type model.

Bjørnstad and Falck (2001) propose a nonparametric estimate of the spatial covariance matrix moving from a continuous nonparametric positive semidefinite estimator of $f(d_{ij}, \phi)$ in (18), called spline correlogram. In particular, they build on the seminal work of Hall and Patil (1994) who, in turn, develop a kernel estimator of the spatial autocorrelation function $\rho(s_i, s_j)$:

$$\tilde{\rho}(s_i, s_j) = \frac{\sum_{i=1}^n \sum_{j=1}^n K(d_{ij}/a)(\hat{\rho}_{ij})}{\sum_{i=1}^n \sum_{j=1}^n K(d_{ij}/a)} \quad (19)$$

where K is a kernel function, a is a bandwidth and $\hat{\rho}_{ij}$ is the sample correlation

$$\hat{\rho}_{ij} = \frac{(z_i - \bar{z})(z_j - \bar{z})}{1/n \sum_{l=1}^n (z_l - \bar{z})^2} \quad (20)$$

in which $\bar{z} = 1/n \sum_{l=1}^n z_l$ is the sample mean. Hall and Patil (1994) demonstrate that the estimator in (19) can be tuned (by tuning a) so that $\tilde{\rho}(d) \rightarrow 0$ for any smooth functional form of $\rho(d)$.

Starting from the estimator in (19), Bjørnstad and Falck (2001) opt for a cubic B-spline as a smoother⁷. Given N pairs (X_i, Y_i) , $i = 1, \dots, N$, the smoothing spline solves the fitting problem by selecting the function f that minimizes the penalized residual sum of squares (RSS)

$$RSS(f(X), \tau) = \sum_{i=1}^N \{Y_i - f(X_i)\}^2 + \tau \int \{f''(t)\}^2 dt \quad (21)$$

where τ is a fixed smoothing parameter (Hastie *et al.*, 2009). In the above expression, the first term measures closeness to the data, the second term penalizes curvature in the function; τ represents a trade-off between the two, varying from very rough fits ($\tau = 0$) to very smooth fits ($\tau = \infty$). The asymptotic kernel, equivalent to a cubic B-spline is:

$$K(d/a) = \frac{1}{2} \exp\left(-\frac{|d/a|}{\sqrt{2}}\right) \sin\left(-\frac{|d/a|}{\sqrt{2}} + \frac{\pi}{4}\right) \quad (22)$$

where, once more, d denotes a generic measure of distance. The advantage in using the B-spline is in that this smoother adapts better to irregularly spaced data and produces a

⁷ Silverman (1984) points out that the smoothing spline is essentially a local kernel average with a variable bandwidth.

consistent estimate of the covariance function (Hyndman and Wand, 1997). It has been shown that fixing the degree of smoothing using cross validation (see Green and Silverman, 1994, and Hastie *et al.*, 2009, for more details) and assuming a true covariance function $\rho(s_i, s_j)$ that is C2-differentiable (i.e. with continuous 1st and 2nd derivatives) guarantees results with asymptotic properties.

In addition, since the estimator $\tilde{\rho}(s_i, s_j)$ must be not only consistent but also positive semidefinite, and this is not necessarily guaranteed by the estimator in equation (19), Bjørnstad and Falck resort to a Fourier-filter method (Hall *et al.*, 1994). The latter works as follows: firstly the Fourier transform of $\tilde{\rho}(s_i, s_j)$ is calculated, then all negative excursions of the transformed function are set to zero and, finally, a nonparametric positive semidefinite estimate of the spatial correlation function is obtained by backtransformation.

3.3.2 The SNP estimator in practice

The SNP estimator finds a natural application in estimating the very general model (10), previously introduced. Operatively, the SNP estimator of $M(\mathbf{X})$ is computed through the following steps:

0. *Pilot fit*: estimate $M(X)$ with a local polynomial smoother, where the bandwidth h is chosen following an optimal rule. As for the degree of the polynomial, we we fix it to 1. The output is $\hat{u} = Y - \hat{M}(X)$.
1. *Nonparametric covariance matrix estimation*: using the spline correlogram, obtain \hat{V} , the estimated spatial covariance matrix of \hat{u} .
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} .

3.3.3 Monte Carlo study

Finally, we conduct a Monte Carlo experiment⁸ to show, via simulations, the finite sample performance of our procedure in comparison with a traditional nonparametric method that does not take the presence of spatial dependence into account. The purpose therefore is to investigate the effective improvement in regression estimation results when spatial dependence is not neglected.

The Monte Carlo experiment is carried out considering several nonlinear specifications for model (10)

$$\begin{aligned} Y &= M(X) + u \\ u &= \rho W_y u + \epsilon \end{aligned}$$

In particular, we consider the following specifications that correspond to complex functional forms, capable of incorporating the nonlinearity due to the factor $(I - \rho W_y)^{-1}$:

- 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 $M(X) = 10\exp(-10X)$
- E $M(X) = (-1 + 2X) + 0.95\exp(-40(-1 + 2X)^2)$
- F $M(X) = 1/(1 + \exp(-6 + 12X))$
- G $M(x) = \begin{cases} \exp(X - 0.33) & \text{if } X < 0.33 \\ \exp(-2(X - 0.33)) & \text{if } X \geq 0.33 \end{cases}$

the shapes of which 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,

⁸ The core of the code has been written in Matlab (Matlab 7.7.0, R2008b), but it incorporates some R (R Development Core Team, 2010) functions. Connectivity between Matlab and R is ensured via the MatlabRlink toolbox and StatConnDCOM (Baier and Neuwirth, 2007).

$X \sim U(0, 1)$, while the disturbance term is generated as a vector of normally distributed random variables, $\epsilon \sim N(0, \sigma^2)$, where σ is set to obtain three alternative levels for the pseudo- R^2 (0.2, 0.5, 0.8). In addition, units are assumed to belong to a circular world and, similar to Kelejian and Prucha (1999 and 2007) and Kapoor *et al.* (2007), the spatial weights matrix W is such that each observation is directly related to the six units immediately surrounding it (three on each side) it in the ordering. Specifically, the matrix is such that all nonzero elements are equal and, following common practice, the matrix is row-normalized. Finally, ρ takes on three alternative values (0.3, 0.5, 0.8) corresponding to low, intermediate and strong spatial dependence, giving us a total of 189 ($7 \times 3 \times 3 \times 3$) experiments.

We employ two estimation methods: the traditional local linear estimator (NP) and the procedure proposed in the previous Section (SNP), implemented with a local linear estimator. For all simulations we use the gaussian kernel with bandwidths that minimize the cross-validation criterion.⁹ The starting value for the smoothing parameter in the estimate of the spline correlogram is 0.1.¹⁰

An estimator's performance is measured by calculating the median across replications of the Mean Integrated Squared Error (MISE) obtained in each replication. 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 on NP. The results of the complete set of experiments are reported in Table 1.

Overall, the performance of SNP is quite good as median ratios are in almost all cases below 1, with no appreciable differences across the different functional forms. Median ratios are close to 1 for low levels of spatial dependence ($\lambda = 0.3$) 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, SNP procedure visibly outperforms the

⁹ 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 optimal bandwidth obtained via the cross-validators criterion.

¹⁰ Operatively, the estimate of the spatial autocorrelation function is obtained through two subsequent smoothings and the corresponding smoothing parameters are chosen, respectively, by the user and then by generalized cross-validation minimization

traditional local linear estimator when ρ reaches 0.8 and $N = 200$, obtaining median values of the MISE that are approximately 25 percent smaller in several cases.

4 Empirical analysis

We can now apply the ideas developed in the previous sections to explain observed patterns of convergence among US states between 1971:Q1 and 2010:Q4. As shown in Magrini *et al.* (2015) and Gerolimetto and Magrini (2014), when regional disparities follow a distinct cyclical pattern in the short-run, results from convergence analysis could be affected by sizeable distortions if the analyzed period includes incomplete cycles. In order to avoid this bias, we resort to the approach in Gerolimetto and Magrini (2014) and begin the analysis by extracting the trend from each of the 48 quarterly per capita personal income series using the Hodrick-Prescott filter (Hodrick and Prescott, 1997). The choice of λ_{HP} , the parameter that controls the degree of smoothness of the estimated trend, is made following the logic spelt out in Gerolimetto and Magrini (2014) and is set equal to 10000.

Once the trends have been estimated, we fix two points in time, t and $t+s$, and study convergence dynamics over this specific time period by applying the distribution dynamics approach to data on the extracted trends. For each considered time period, we estimate the stochastic kernel and, when possible, we calculate the corresponding ergodic distribution i.e., a 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 that corresponds to a given stochastic kernel by solving

$$f_{\infty}(Y) = \int_{-\infty}^{\infty} f(Y_{t+s}|Y_t) f_{\infty}(Y) dY \quad (23)$$

Then, we compare stochastic kernels and corresponding ergodic distributions estimated using both the traditional local linear estimator (NP), and our estimator (SNP) in the mean's function adjustment procedure.¹¹ To get an idea of the speed with which the distributions

¹¹ When using SNP, the spatial correlogram is estimated using a matrix of orthodromic distances between

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 10-nearest neighbors, row-standardized matrix) on observed and filtered data as well as on the residuals of the regression for the mean function estimation, and two dispersion measures (coefficient of variation and interquartile range) for initial, final and ergodic distributions. Finally, we also study the way the cross-sections distribution evolves towards the steady-state with the specific aim of figuring out which units contribute the most to the shape of the tails of the ergodic distribution. In particular, based on SNP estimates, we plot the conditional distribution for each state’s initial per capita personal income level after a number of iterations corresponding to the half-life of the chain. Then, we identify the conditional distributions (and hence the states) for which either the third quartile is smaller than 1 or the first quartile is larger than one. In such a way, we are able to recognize which states are more likely to end up in one of the tails during the transition towards the ergodic distribution.

Drawing directly from the results in Gerolimetto and Magrini (2014), we initially concentrate on two sub-periods stretching, respectively, between 1971:Q1 and 1980:Q4 and between 1981:Q1 and 2010:Q4. As far as the first sub-period is concerned, Figure 2 displays a clear result of persistence, which is common to both estimators despite the presence of strong spatial correlation not only in the data but also in the residuals from the NP regression (Table 2, columns 1-2). In other words, in this sub-period, allowing for spatial dependence does not alter the end result in any significant way.

A somewhat different outcome instead emerges from the analysis of the second sub-period as, both the plots in Figure 3 and of the columns 3-4 of Table 2, suggest that the estimates obtained using SNP appear characterized by a weaker tendency towards divergence compared to NP. To ascertain the significance of the differences in the two ergodic state capitals.

distributions, we carry out the Cramér-von Mises test for the equality of two continuous distributions. As shown in Table 4, the null hypotheses that the two samples are from the same continuous distribution is rejected at all significance levels. Hence, allowing for spatial dependence significantly modifies the estimation results over this sub-period and, in particular, the estimates through SNP depict a weaker tendency towards divergence than what originally suggested using NP.

Given these statistically significant differences, it appears important to split the second sub-period into smaller ones to investigate whether the difference in outcome is confined to specific decades. Starting from the 1981:Q1 and 1990:Q4 sub-period, it is interesting to note that the ergodic distribution cannot be calculated according to (23) when NP is employed. The shape of the estimated stochastic kernel (Figure 4, right column), however, clearly suggests a strong tendency towards divergence given the pronounced counter-clockwise twist of the conditional probability in correspondence of values above 1.20. A rather different result is found when SNP is employed: not only the ergodic distribution can now be calculated but its shape and the statistics reported in Table 2, columns 5-6, suggest that this period is characterised by a tendency towards persistence. However, it is also interesting to note that, compared to the initial distribution, the ergodic exhibits a heavier left tail and a lighter right one.

When the 1991:Q1 and 2000:Q4 sub-period is considered, the external shape of the initial and final distributions is remarkably similar as the corresponding plots almost overlaps. There are however intra-distribution movements and these lead to ergodic distributions that suggest a clear tendency towards divergence. Two aspects of this process are worth emphasizing. First of all, this divergence process is very slow as witnessed by the high half-life values (Table 3); for instance, the half-life of the estimate obtained via SNP reaches the value of 8.5, a 5-fold increase with respect to the value corresponding to the period 1981:Q1-1990:Q4. Secondly, although all estimates suggest a process of divergence (Figure 5 and Table 2, columns 7-8), the tendency towards divergence appears stronger when SNP

is employed. Interestingly, the shape of the estimated stochastic kernel and that of the ergodic distribution obtained using SNP suggest that this decade is characterised by a process of club-convergence. In fact, there appears to be a difference in the strength of the diverging process highlighted by the two estimators since the null hypothesis is only marginally accepted as shown in Table 4.

The last sub-period (2001:Q1 and 2010:Q4) seems to delineate a strong tendency towards divergence both when NP and SNP are employed in the mean bias adjustment procedure (Figure 6 and Table 2, columns 9-10). In particular, the tests for the equality of the distributions suggest no statistically significant differences between the two ergodic distributions (Table 4).

Finally, Table 5 illustrates the transitional dynamics.¹² In the first sub-period (1971:Q1-1980:Q4), we note that Alabama, Arkansas, Kentucky, Louisiana, Mississippi, South Carolina and West Virginia are more likely to remain in the lower left tail of the distribution while California, Connecticut, Delaware, Illinois, Maryland, Nevada, New Jersey, New Mexico and New York are instead more likely to stay in the upper right tail.

In the second sub-period (1981:Q1-1990:Q4), only units on the lower left tail of the distribution can be identified; these are Alabama, Arkansas, Kentucky, Maine, Mississippi, South Carolina, Tennessee and West Virginia. These is clearly consistent with the shift of the probability mass towards the left tail noted in the corresponding ergodic distribution noted in the right column of Figure 4.

Contrary to the previous case, in the third sub-period (1991:Q1-2000:Q4), only units on the upper right tail of the distribution can be identified (Connecticut, Maryland, Massachusetts, New Jersey and New York). In addition, for the states of Maryland, Massachusetts and New York, the mode of the conditional distribution is higher than the level of per capita income that these states have at the beginning of the sub-period. In other words, these states are likely to diverge from the rest of the distribution and are likely to

¹²The transitional dynamics of each state in each sub-period are depicted in Figures A1-A8 in the Appendix.

contribute relevantly to the overall divergence result conveyed by the shape of the ergodic distribution in Figure 5.

Finally, a similar picture emerges in the last sub-period (2001:Q1-2010:Q4), in which Colorado, Connecticut, Maryland, Massachusetts and New Jersey are likely to remain in the upper right tail of the distribution but only two of them (Connecticut and New Jersey) show a strongly diverging pattern.

5 Conclusions

In this paper we studied the evolution of per capita personal income inequalities among US states using data from the 1st quarter of 1969 to the 4th quarter of 2012. Specifically, we employed a distribution dynamics approach in which the conditional mean is estimated via a nonparametric estimator that allows for spatial dependence (SNP) and compared its results with those obtained from a conventional nonparametric estimator (NP).

Comparing estimated conditional densities and ergodic distributions we found particularly interesting differences in the results obtained over the three decades running between 1981:Q1 and 2010:Q4. With reference to the first decade (1981:Q1 and 1990:Q4), by employing the SNP estimator we found clear evidence of persistence, contrary to findings of strong divergence obtained using NP. In the second decade (1991:Q1 and 2000:Q4), the SNP estimator allowed us to detect a stronger tendency towards divergence, possibly leading to club convergence. Instead, both estimators allowed to detect statistically significant evidence of divergence during the last decade (2001:Q1 and 2010:Q4).

From a more general viewpoint, we emphasize one important conclusion that stems from this analysis: not only we confirm that neglecting spatial dependence might substantially affect the results, but we also suggest that the direction in which spatial dependence alters the results is not easily predictable.

References

- Ansari, A.R., Bradley R.A. (1960): Rank-sum tests for dispersions. *Annals of Mathematical Statistics*, **31**, 1174–1189
- Azzalini, A., Bowman A.W. (1997): Applied Smoothing Techniques for Data Analysis. Clarendon Press, Oxford
- Baier. T., Neuwirth, E. (2007): Excel::COM::R. *Computational Statistics*, **22**, 91–108
- Bartels, C.P., Ketellapper, R. (1979): Exploratory and Explanatory Analysis of Spatial Data. Martinus Nijhoff, Boston.
- 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, London.
- Bjørnstad, O.N., Falck, W.(2001): Nonparametric Spatial Covariance Functions: Estimation and Testing. *Environmental and Ecological Statistics*, **8**, 53–70
- Cliff, A.D. and Ord, J.K. (1973): Spatial Autocorrelation. London: Pion.
- Durlauf, S.N., Johnson, P.A., Temple, J.R.W.(2005): Growth Econometrics. In: Aghion, P., Durlauf, S.N. (eds.) Handbook of Economic Growth, Volume 1. Elsevier, Amsterdam
- Durlauf, S.N., Quah, D.T.(1999): The New Empirics of Economic Growth. In: Taylor, J., Woodford, M. (eds.) Handbook of Macroeconomics. Elsevier, Amsterdam
- 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**, 109–140

- Florax, R., Folmer, H. (1992) Specification and estimation of spatial linear regression models. Monte Carlo evaluation of pre-test estimators. *Regional Science and Urban Economics*, **22**, 405–432
- Gerolimetto, M., Magrini, S. (2014): A novel look at long-run convergence dynamics in the US. *International Regional Science Review*, Forthcoming
- Getis, A., Ord. J.K. (1992): The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, **24**, 189-206
- Green, P.J., Silverman, B.W. (1994): *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. London: Chapman & Hall
- Hall, P., Patil, P. (1994): Properties of Nonparametric Estimators of Autocovariance for Stationary Random Fields. *Probability Theory and Related Fields*, **99**, 399–424
- Hall, P., York, M. (2001): On the calibration of Silvermans test for multimodality *Statistica Sinica*, **11**, 515–536.
- Hall, P., Fisher, N.I., Hoffman, B. (1994): On the Nonparametric Estimation of Covariance Functions. *Annals of Statistics*, **22**, 2115–2134.
- Hastie, T., Tibshirani R., Friedman J. (2009): *The Elements of Statistical Learning*. Berlin: Springer.
- Hodrick, R., E.C. Prescott. (1997): Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit and Banking*, **29**, 1–16
- Hyndman, R.J., Wand, M.P. (1997): Nonparametric Autocovariance Function Estimation. *Australian Journal of Statistics*, **39**, 313–324
- Hyndman, R.J., Bashtannyk, D.M., Grunwald, G.K.(1996): Estimating and Visualizing Conditional Densities. *Journal of Computational and Graphical Statistics*, **5**, 315–336

- Islam (2003): What Have We Learnt from the Convergence Debate? *Journal of Economic Surveys*, **17**, 309–362
- Jenish, N., Prucha I. (2009): Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics*, **150**, 86–98
- Johnson, P.A. (2005): A Continuous State Space Approach to Convergence by Parts. *Economic Letters*, **86**, 317–321.
- Kapoor, M., Kelejan H.H., Prucha I. (2007): Panel data models with spatially correlated error components. *Journal of Econometrics*, **140**, 97-130
- Kelejian, H.H., Prucha, I.R. (1999): A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic Review*, **40**, 509–533
- Kelejian, H.H., Prucha, I.R. (2007): HAC Estimation In a Spatial Framework. *Journal of Econometrics*, **140**, 131–154
- Lin, X., 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
- Martins-Filho, C., Yao, F. (2009): Nonparametric Regression Estimation with General Parametric Error Covariance. *Journal of Multivariate Analysis*, **100**, 309–333
- Magrini, S.(2004): Regional (Di)Convergence. In: Henderson, J.V., Thisse J.F. (eds.) Handbook of Regional and Urban Economics, Volume 4. Elsevier, Amsterdam
- Magrini, S.(2009): Why Should We Analyse Convergence Using the Distribution Dynamics Approach? *Scienze Regionali - Italian Journal of Regional Science*, **8**, 5–34

- Magrini, S., Gerolimetto, M. and H.E. Duran. (2015) Regional Convergence and Aggregate Business Cycle in the US. *Regional Studies*, **49**, 251–272
- Maza, A., Hierro, M. and J. Villaverde. (2010) Measuring Intra-Distribution Dynamics: An Application of Different Approaches to the European Regions. *Annals of Regional Science*, **45**, 313–329
- Nadaraya, E.A. (1964): On Estimating Regression. *Theory of Probability and its Applications*, **9**, 141–142
- Nadaraya, E.A. (1964): On Estimating Regression. *Theory of Probability and its Applications*, **9**, 141–142
- Paelinck, J., Klaassen, L. (1979): Spatial Econometrics. Farnborough: Saxon House.
- 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
- R Development Core Team (2010): R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

- Rey, S.J., Montouri, B.D. (1999): US Regional Income Convergence: A Spatial Econometric Perspective. *Regional Studies*, **33**, 143-156
- Robinson, P.M. (1987): Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form. *Econometrica*, **55**, 875–891
- Robinson, P.M. (2008): Developments in the Analysis of Spatial Data. *Journal of the Japanese Statistical Society*, **38**, 87–96
- Robinson, P.M. (2011): Asymptotic Theory for Nonparametric Regression with Spatial data. *Journal of Econometrics*, **165**, 5–19
- 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
- Shorrocks, A.F. (1978): The Measurement of Mobility. *Econometrica*, **46**, 1013–1024
- Silverman, B.W. (1981): Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society, Series B* **43**, 97–99.
- Silverman B.W. (1984): Spline Smoothing: The Equivalent Variable Kernel Method. *Annals of Statistics*, **12**, 898–916
- Wang, N. (2003): Marginal Nonparametric Kernel Regression Accounting for within Subject Correlation. *Biometrika*, **90**, 43–52
- Watson, G.S. (1964): Smooth Regression Analysis. *Sankhya, Series A*, **26**, 359–372
- 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

Figures and Tables

Figure 1: Monte Carlo experiment: functional forms

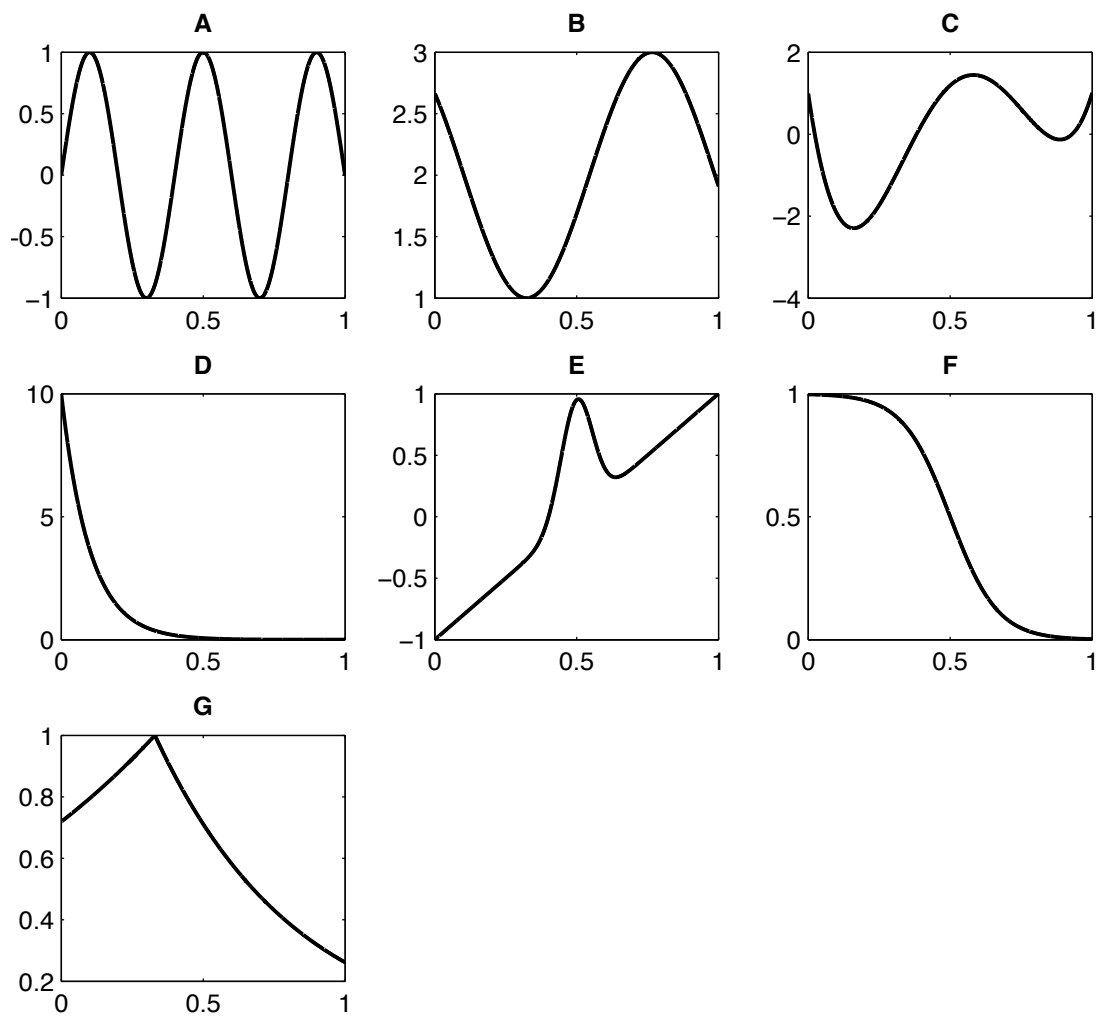
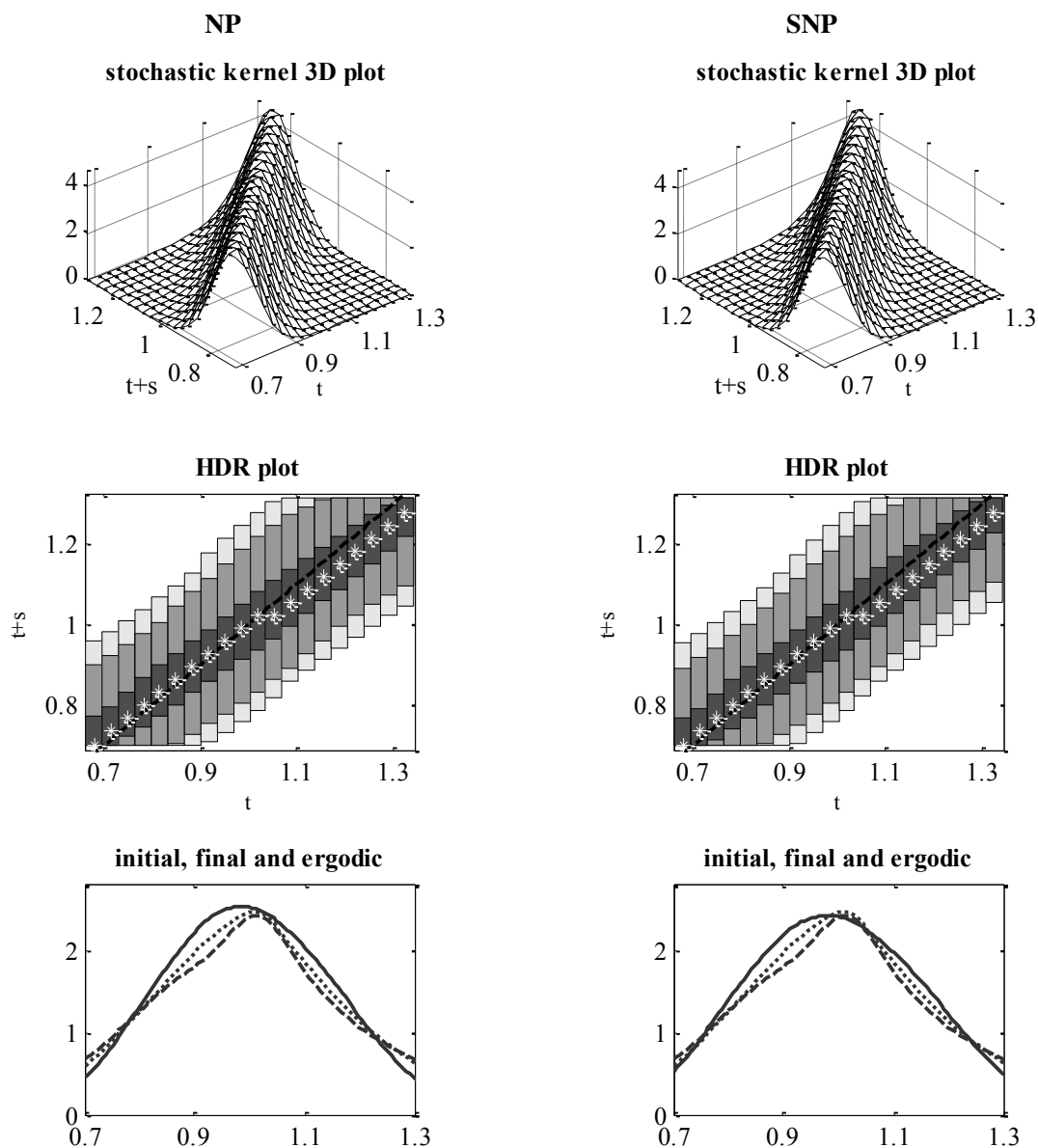
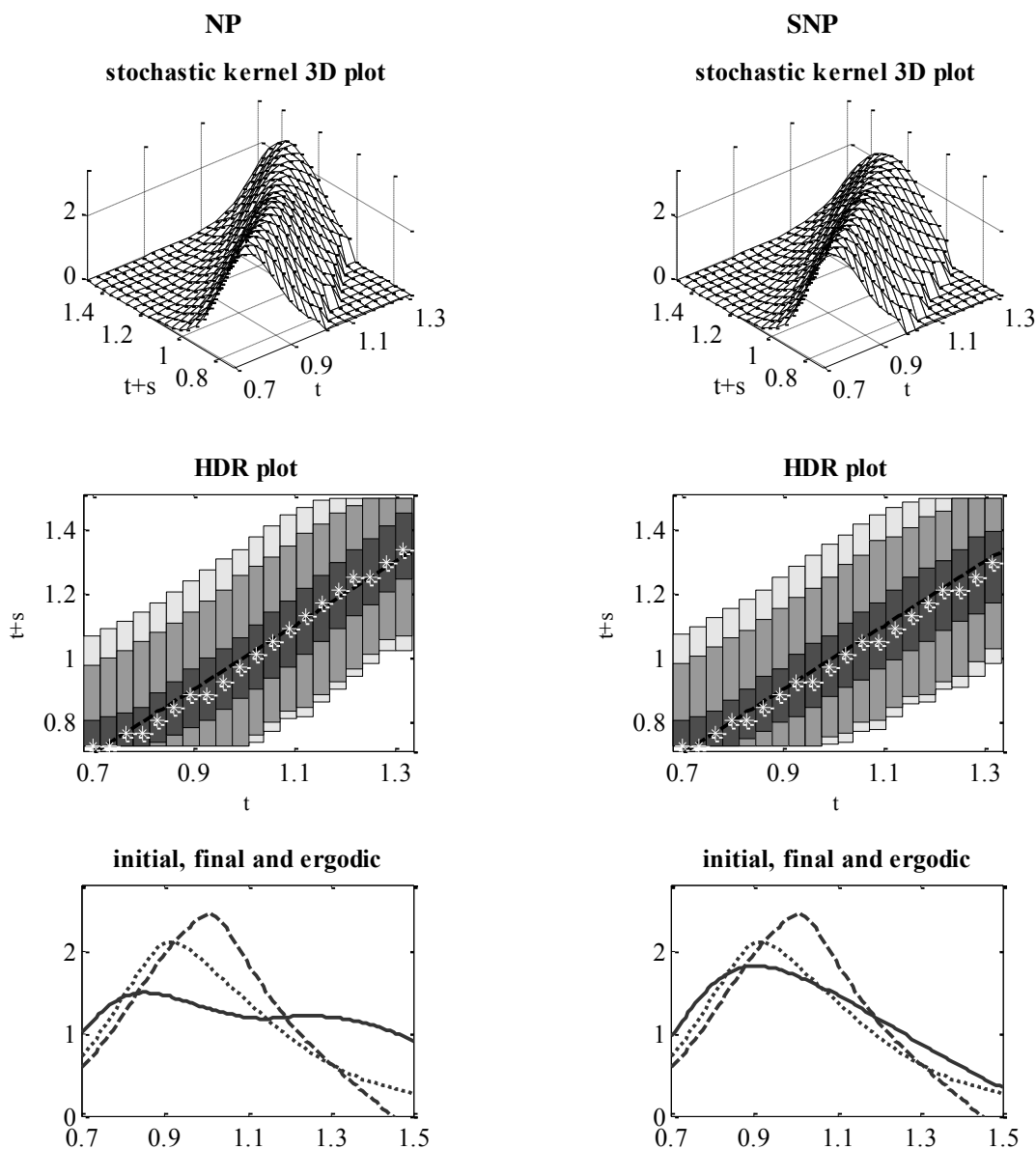


Figure 2: 1971:Q1-1980:Q4



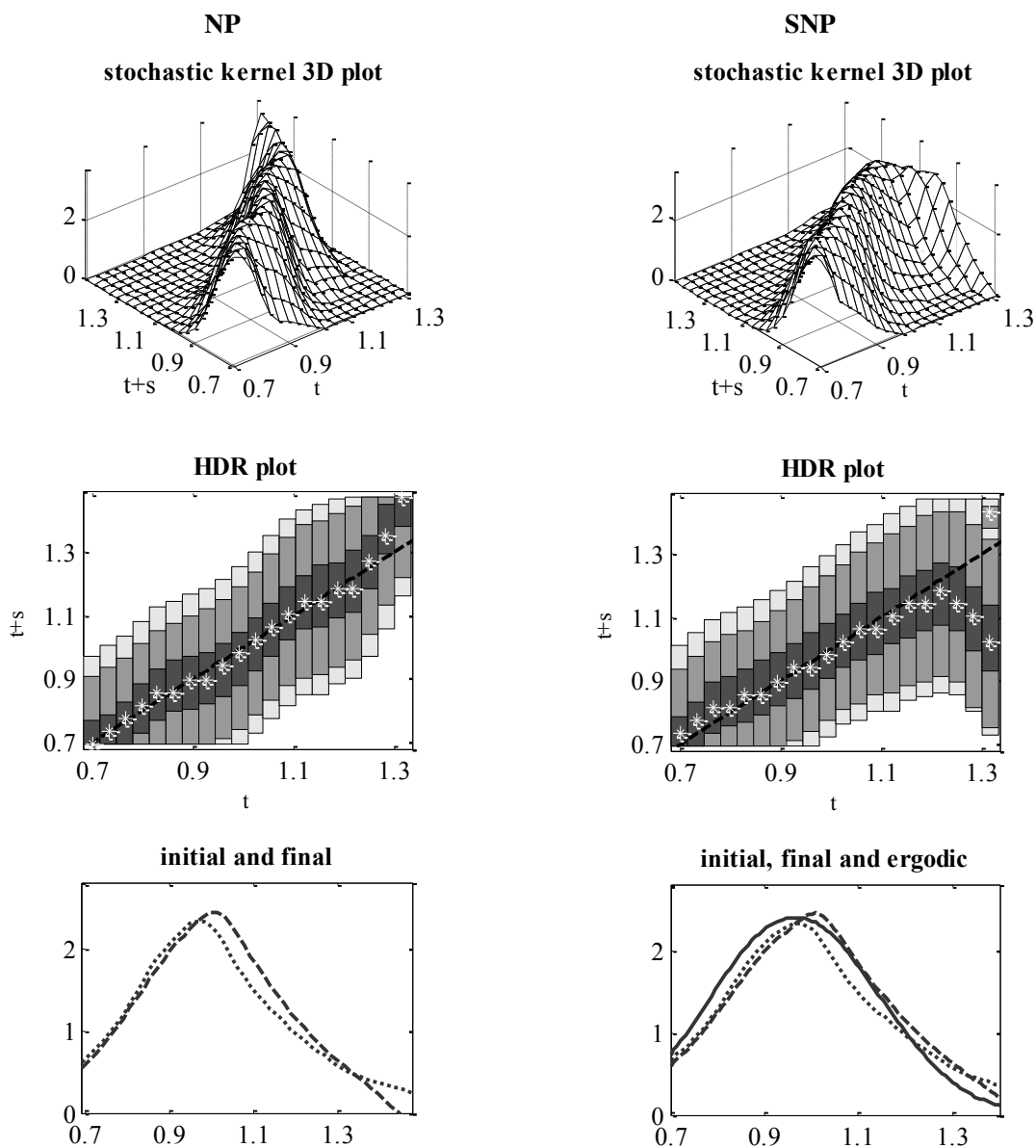
Notes: Estimates of the stochastic kernel use a nearest-neighbor bandwidth in the initial year dimension (span = 0.3), a Normal Scale (Silverman, 1986) bandwidth in the final year dimension and a Gaussian kernel. The (fixed) bandwidth for both NP and SNP estimates is chosen using cross validation. The starting value for the smoothing parameter in the estimate of the spline correlogram is 1; the final value is determined via cross-validation minimization. HP-filtered data are obtained setting $\lambda_{HP} = 10000$. In contour and HDR plots, the dashed line represents the main diagonal, the asterisk the modes. In the comparison between distributions, the dashed line represents the initial year, the dotted line represents the final year, the continuous line represents the ergodic.

Figure 3: 1981:Q1-2010:Q4



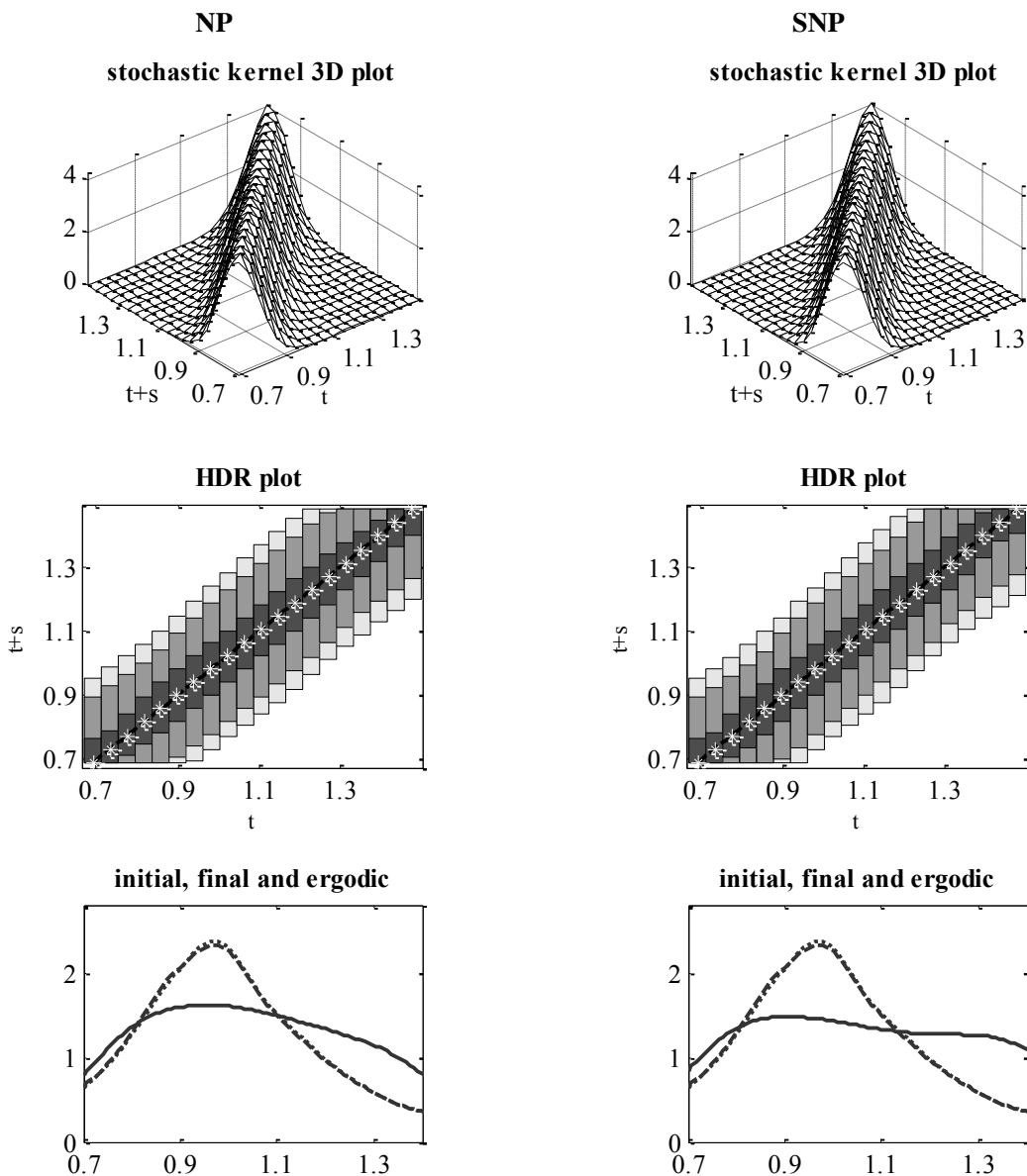
Notes: Estimates of the stochastic kernel use a nearest-neighbor bandwidth in the initial year dimension (span = 0.3), a Normal Scale (Silverman, 1986) bandwidth in the final year dimension and a Gaussian kernel. The (fixed) bandwidth for both NP and SNP estimates is chosen using cross validation. The starting value for the smoothing parameter in the estimate of the spline correlogram is 1; the final value is determined via cross-validation minimization. HP-filtered data are obtained setting $\lambda_{HP} = 10000$. In contour and HDR plots, the dashed line represents the main diagonal, the asterisk the modes. In the comparison between distributions, the dashed line represents the initial year, the dotted line represents the final year, the continuous line represents the ergodic.

Figure 4: 1981:Q1-1990:Q4



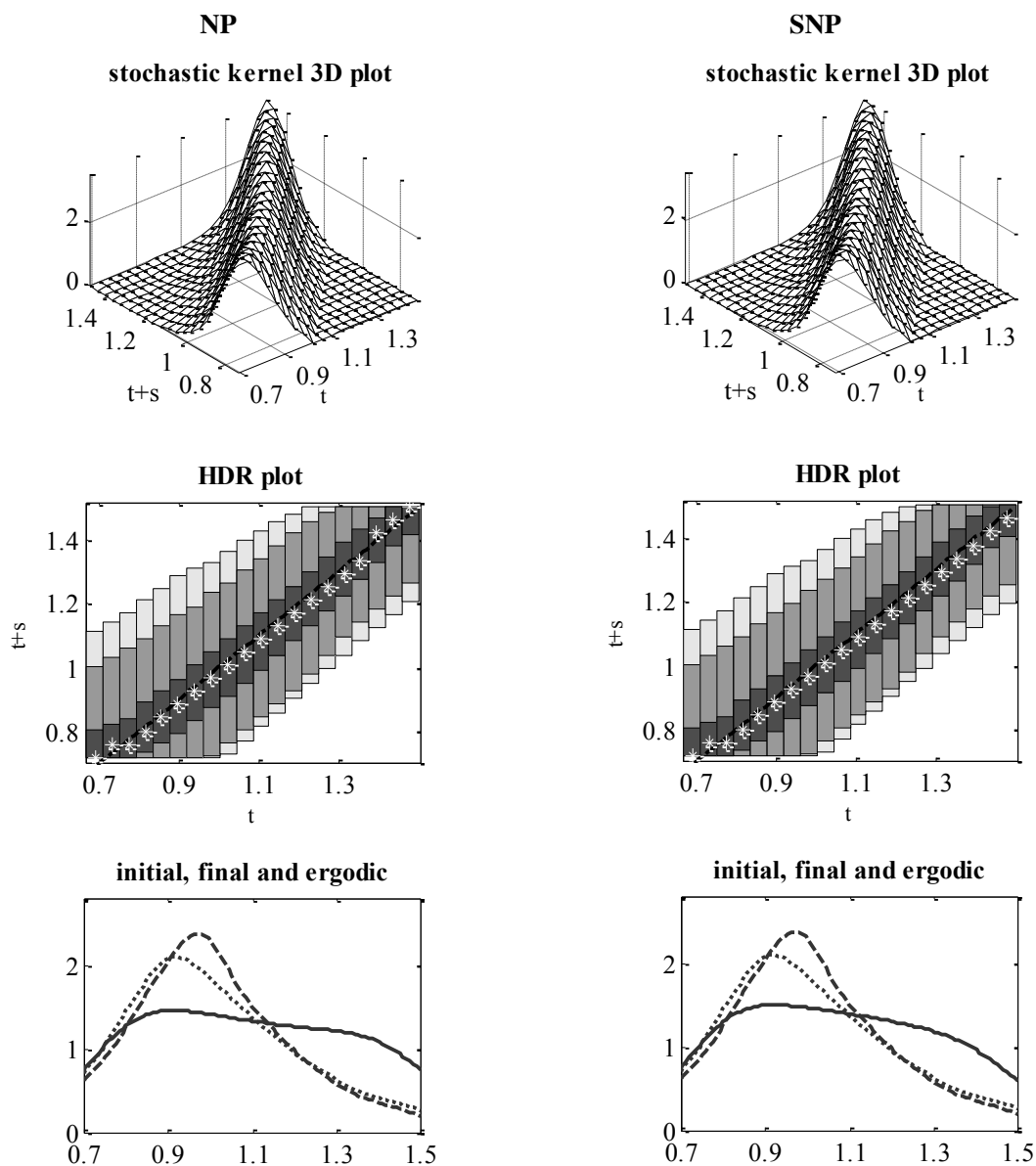
Notes: Estimates of the stochastic kernel use a nearest-neighbor bandwidth in the initial year dimension (span = 0.3), a Normal Scale (Silverman, 1986) bandwidth in the final year dimension and a Gaussian kernel. The (fixed) bandwidth for both NP and SNP estimates is chosen using cross validation. The starting value for the smoothing parameter in the estimate of the spline correlogram is 1; the final value is determined via cross-validation minimization. HP-filtered data are obtained setting $\lambda_{HP} = 10000$. In contour and HDR plots, the dashed line represents the main diagonal, the asterisk the modes. In the comparison between distributions, the dashed line represents the initial year, the dotted line represents the final year, the continuous line represents the ergodic.

Figure 5: 1991:Q1-2000:Q4



Notes: Estimates of the stochastic kernel use a nearest-neighbor bandwidth in the initial year dimension (span = 0.3), a Normal Scale (Silverman, 1986) bandwidth in the final year dimension and a Gaussian kernel. The (fixed) bandwidth for both NP and SNP estimates is chosen using cross validation. The starting value for the smoothing parameter in the estimate of the spline correlogram is 1; the final value is determined via cross-validation minimization. HP-filtered data are obtained setting $\lambda_{HP} = 10000$. In contour and HDR plots, the dashed line represents the main diagonal, the asterisk the modes. In the comparison between distributions, the dashed line represents the initial year, the dotted line represents the final year, the continuous line represents the ergodic.

Figure 6: 2001:Q1-2010:Q4



Notes: Estimates of the stochastic kernel use a nearest-neighbor bandwidth in the initial year dimension (span = 0.3), a Normal Scale (Silverman, 1986) bandwidth in the final year dimension and a Gaussian kernel. The (fixed) bandwidth for both NP and SNP estimates is chosen using cross validation. The starting value for the smoothing parameter in the estimate of the spline correlogram is 1; the final value is determined via cross-validation minimization. HP-filtered data are obtained setting $\lambda_{HP} = 10000$. In contour and HDR plots, the dashed line represents the main diagonal, the asterisk the modes. In the comparison between distributions, the dashed line represents the initial year, the dotted line represents the final year, the continuous line represents the ergodic.

Table 1: Monte Carlo results

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

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

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

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

Table 2: Results

	1971:Q1-1980:Q4		1981:Q1-2010:Q4		1981:Q1-1990:Q4		1991:Q1-2000:Q4		2001:Q1-2010:Q4	
	Moran's I	p-value	Moran's I	p-value	Moran's I	p-value	Moran's I	p-value	Moran's I	p-value
observed initial	0.2606	0	0.1548	0.0011	0.1548	0.0011	0.3141	0	0.251	0
observed final	0.1516	0.0013	0.2854	0	0.3014	0	0.2425	0	0.2854	0
filtered initial	0.2607	0	0.17	0.0004	0.17	0.0004	0.306	0	0.2421	0
filtered final	0.1697	0.0004	0.2865	0	0.3076	0	0.2412	0	0.2865	0
residuals NP	0.0937	0.0280	0.2760	0	0.5215	0	0.2007	0	0.1416	0.0019
residuals SNP	0.0155	0.4815	-0.0840	0.2321	-0.0817	0.2569	0.0410	0.2412	0.0042	0.6264
	CV	IR	CV	IR	CV	IR	CV	IR	CV	IR
filtered initial	0.1563	0.1905	0.1427	0.2144	0.1427	0.1905	0.1677	0.2441	0.1675	0.2440
filtered final	0.1422	0.2077	0.1803	0.2538	0.1681	0.2424	0.1673	0.2458	0.1803	0.2538
ergodic NP	0.1420	0.2036	0.2098	0.3791	-	-	0.1902	0.3158	0.2021	0.3569
ergodic SNP	0.1463	0.2120	0.1914	0.2944	0.1541	0.2136	0.2000	0.3524	0.1989	0.3420

Table 3: Estimated half-life values

	ergodic via SNP	ergodic via NP
1971:Q1-1900Q4	3.5484	3.2583
1981:Q1-2010Q4	2.9073	4.7256
1981:Q1-1990Q4	1.7255	–
1991:Q1-2000Q4	8.4571	7.0202
2001:Q1-2010Q4	4.5453	4.9808

Table 4: Cramér-von Mises test between distributions (test p -levels)

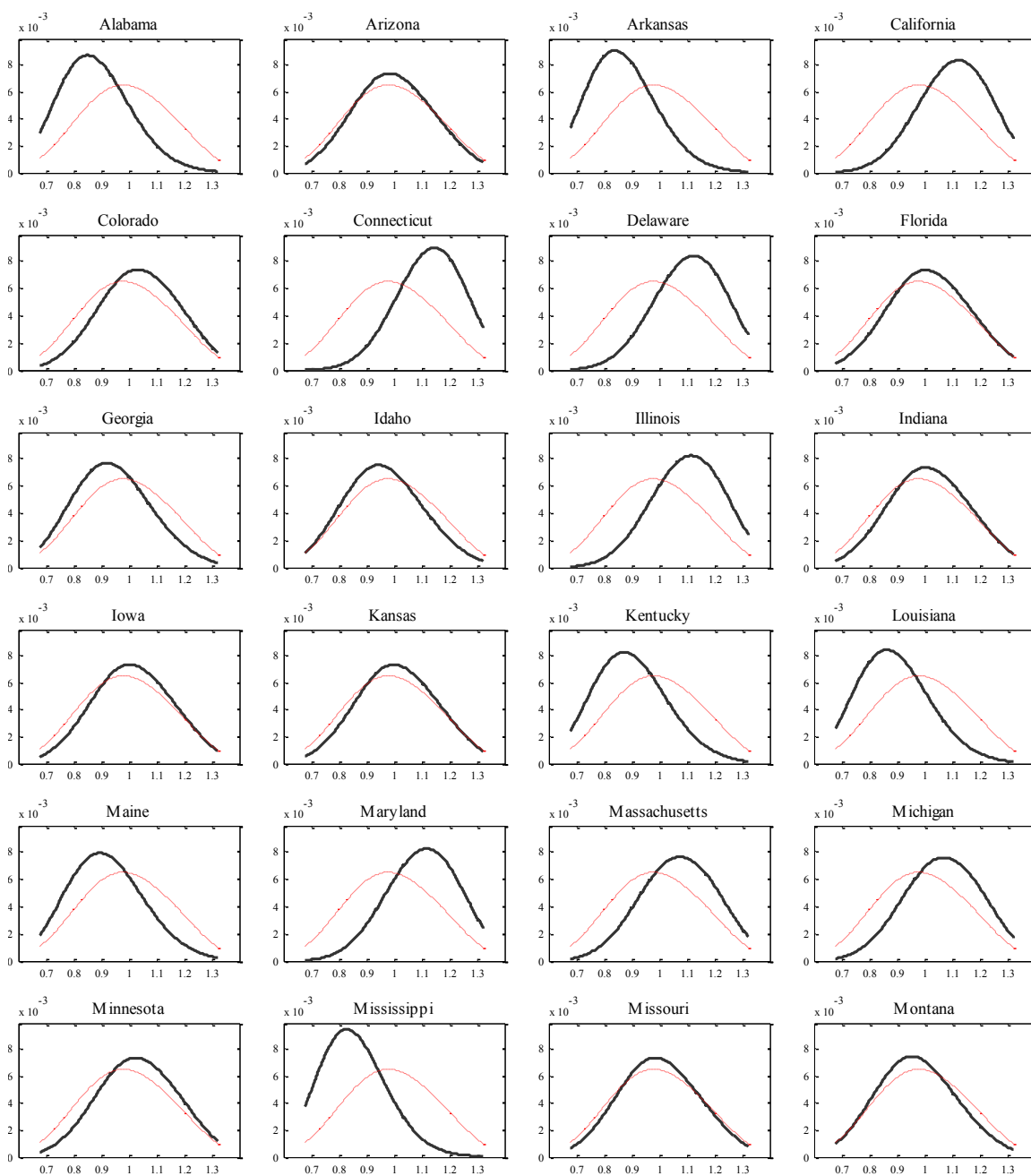
	ergodic SNP vs ergodic NP	ergodic SNP vs initial
1971:Q1-1980Q4	0.9950	0.9736
1981:Q1-2010Q4	0	0
1981:Q1-1990Q4	–	0.0252
1991:Q1-2000Q4	0.1198	0
2001:Q1-2010Q4	0.7545	0

Table 5: Transitional dynamics

	3rd quartile <1	1st quartile >1
1971:Q1-1980:Q4		
Alabama (0.7687)	0.8512	
Arkansas (0.7305)	0.8404	
California (1.2337)		1.1239
Connecticut (1.3236)		1.1482
Delaware (1.2392)		1.1266
Illinois (1.2179)		1.1185
Kentucky (0.8159)	0.8728	
Louisiana (0.798)	0.8647	
Maryland (1.2203)		1.1185
Mississippi (0.6785)	0.8296	
Nevada (1.3006)		1.1428
New Jersey (1.2597)		1.1320
New Mexico (1.2351)		1.1239
New York (1.2351)		1.1239
South Carolina (0.808)	0.8701	
West Virginia (0.7702)	0.8539	
1981:Q1-1990:Q4		
Alabama (0.7985)	0.8590	
Arkansas (0.7641)	0.8459	
Kentucky (0.8237)	0.8656	
Maine (0.8381)	0.8754	
Mississippi (0.6992)	0.7999	
South Carolina (0.8098)	0.8623	
Tennessee (0.8546)	0.8853	
West Virginia (0.7707)	0.8492	
1991:Q1-2000:Q4		
Connecticut (1.4759)		1.3531
Maryland (1.2891)		1.3367
Massachusetts (1.2654)		1.3301
New Jersey (1.3599)		1.3465
New York (1.2399)		1.3202
2001:Q1-2010:Q4		
Colorado (1.217)		1.2093
Connecticut (1.4788)		1.3786
Maryland (1.2577)		1.2702
Massachusetts (1.3152)		1.3244
New Jersey (1.3507)		1.3447

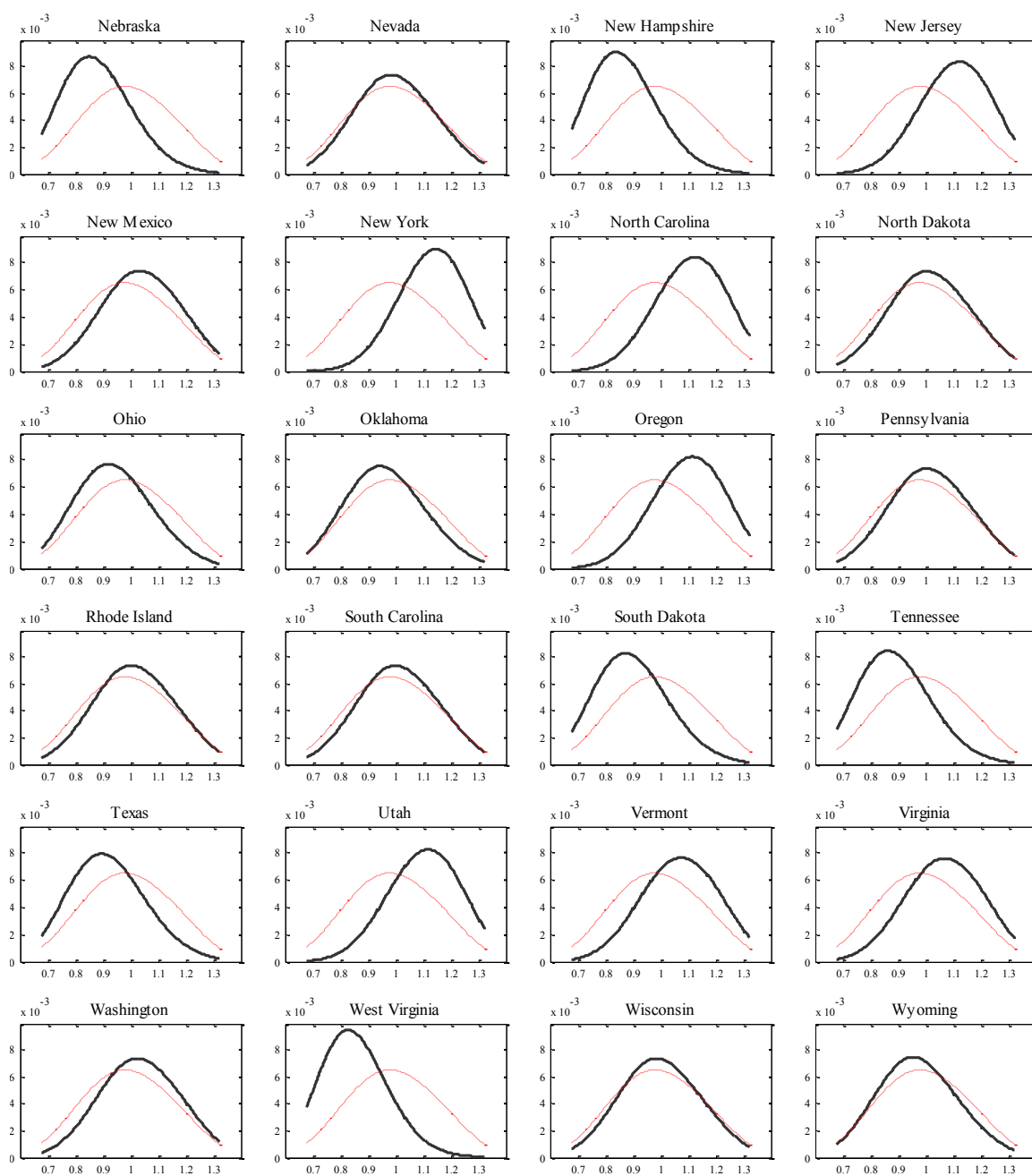
Appendix

Figure A1: Transitional dynamics: 1971:Q1-1980:Q4 (a)



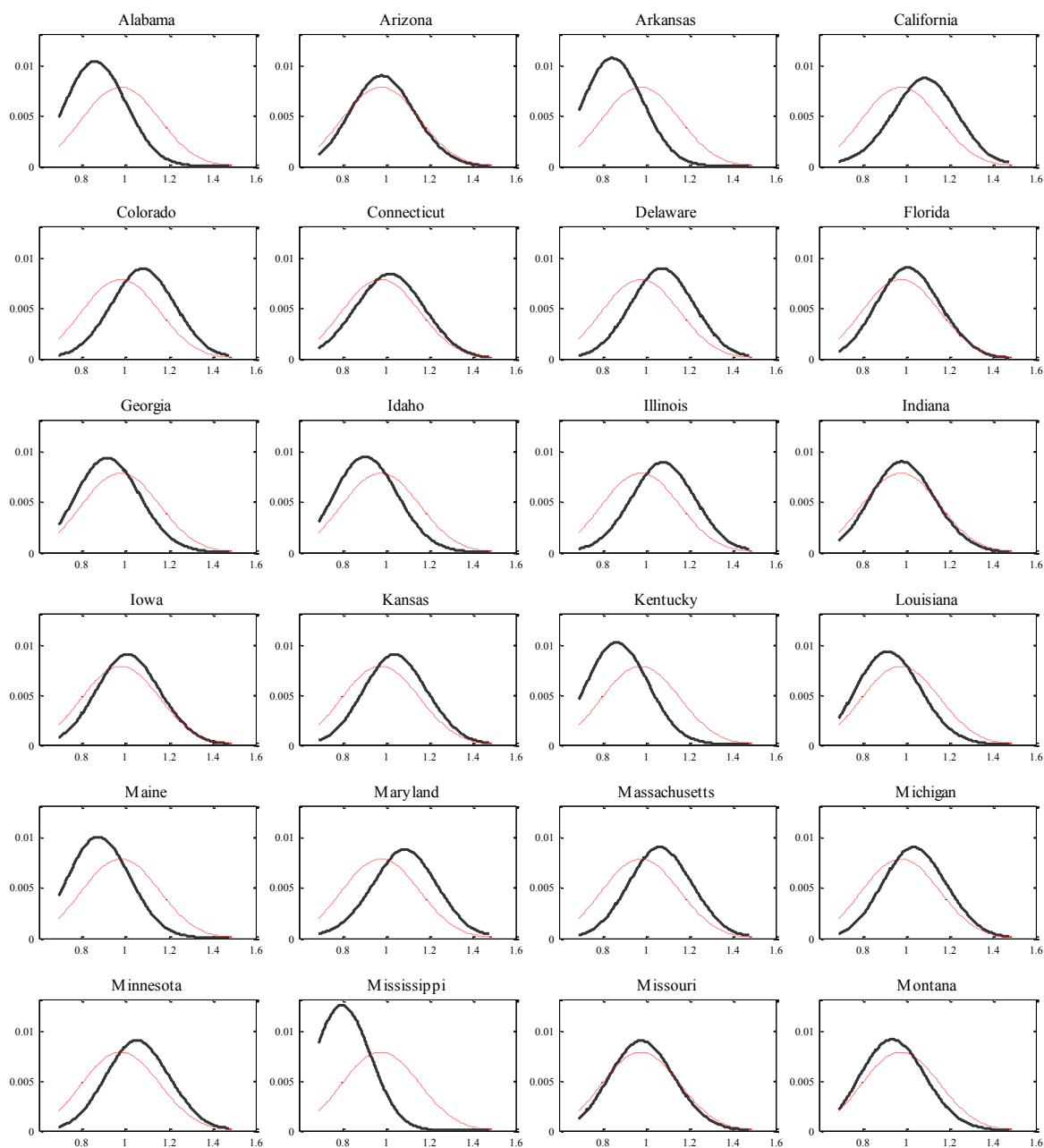
Notes: Plots of the conditional density distributions for each state's initial level of (HP-filtered) per capita income after a number of iterations corresponding to the half-life

Figure A2: Transitional dynamics: 1971:Q1-1980:Q4 (b)



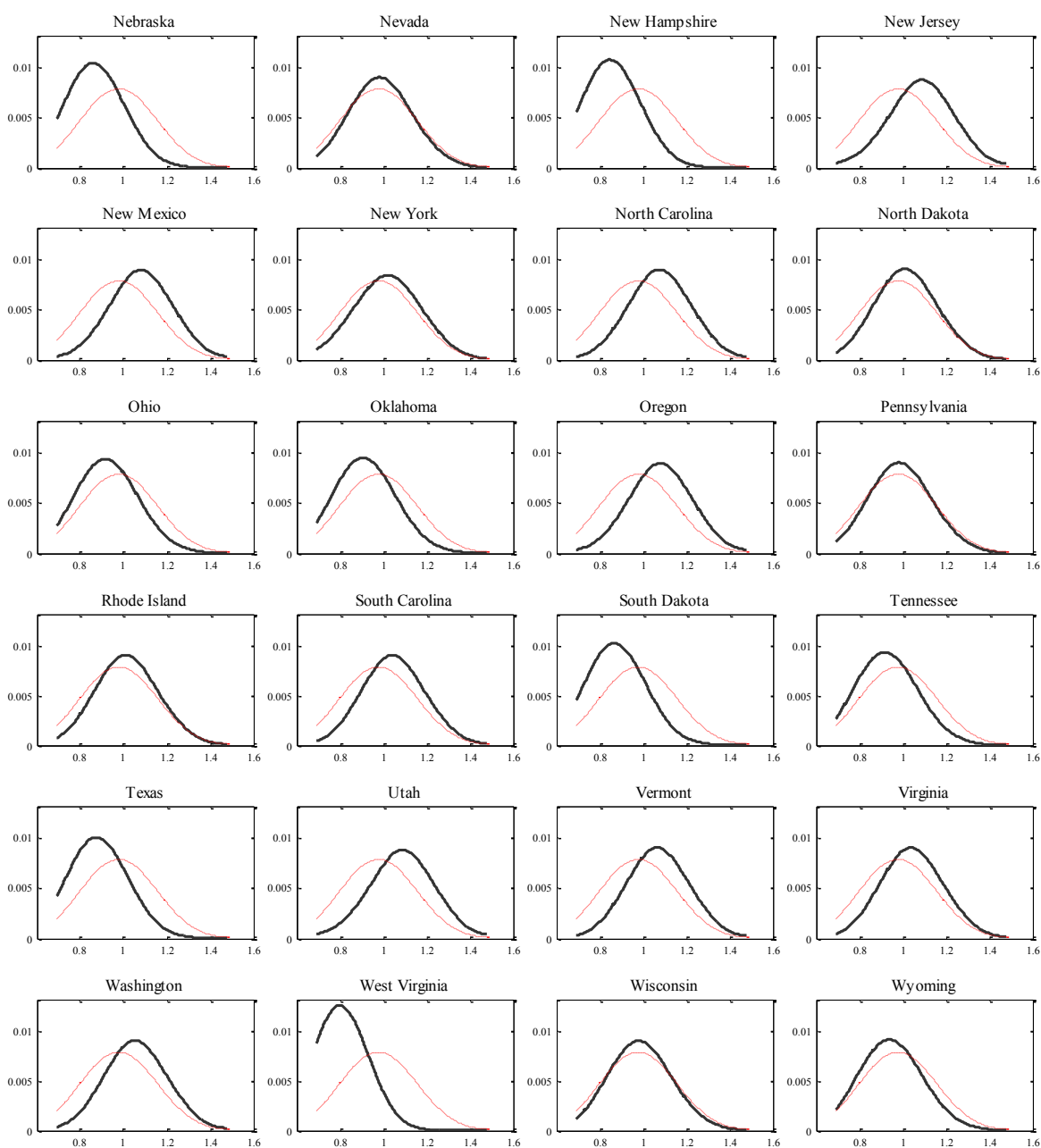
Notes: Plots of the conditional density distributions for each state's initial level of (HP-filtered) per capita personal income after a number of iterations corresponding to the half-life

Figure A3: Transitional dynamics: 1981:Q1-1990:Q4 (a)



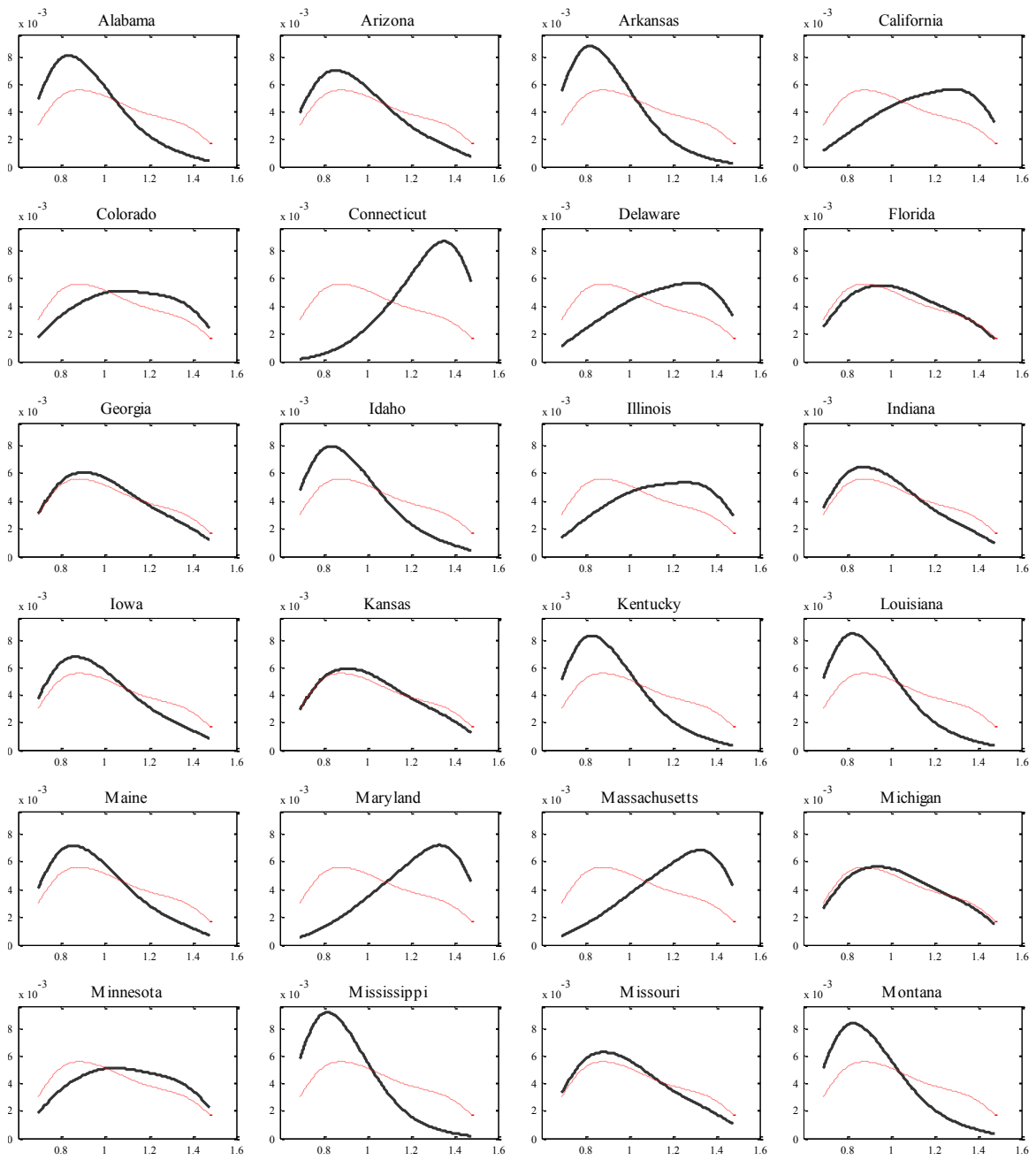
Notes: Plots of the conditional density distributions for each state's initial level of (HP-filtered) per capita personal income after a number of iterations corresponding to the half-life

Figure A4: Transitional dynamics: 1981:Q1-1990:Q4 (b)



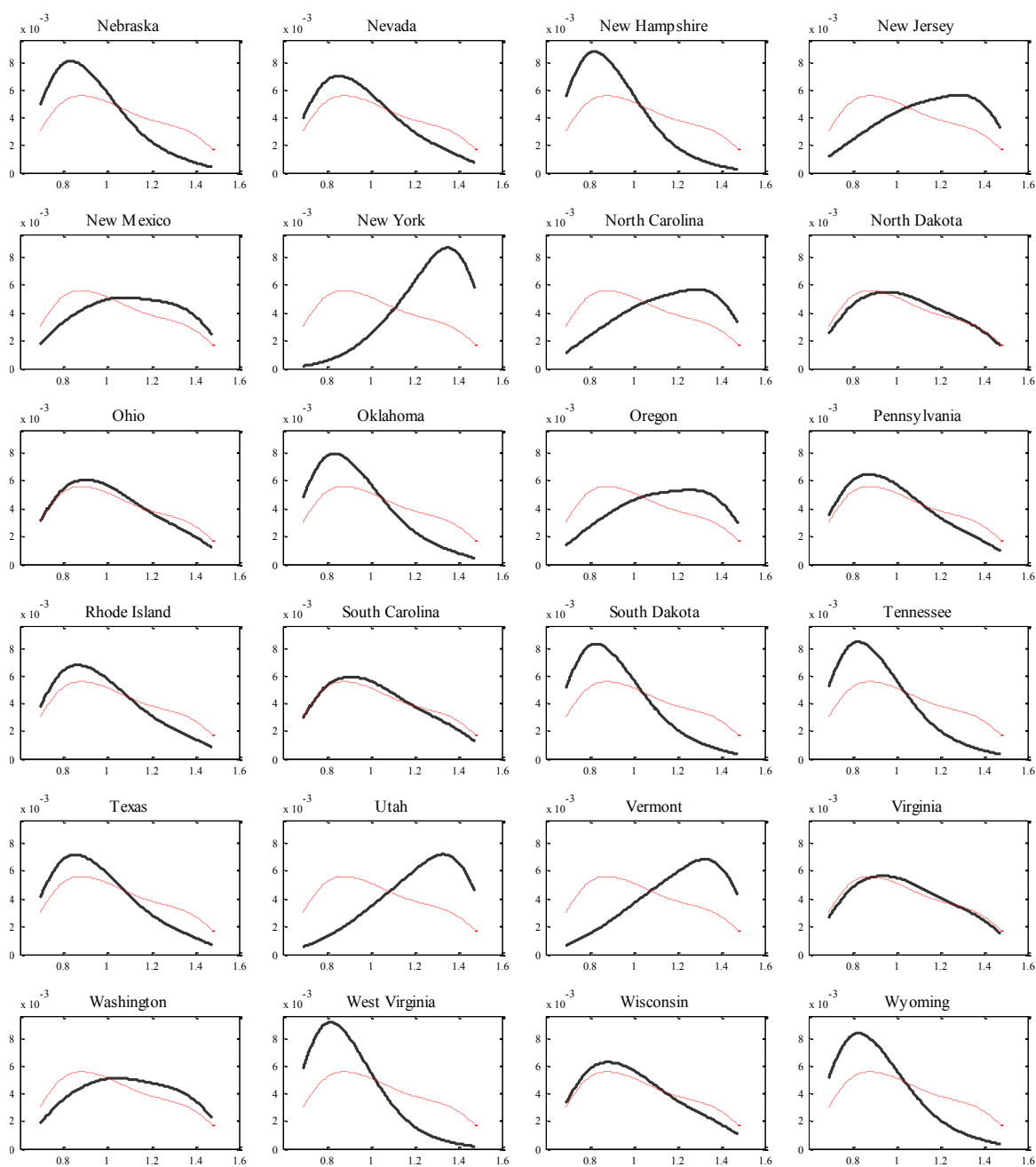
Notes: Plots of the conditional density distributions for each state's initial level of (HP-filtered) per capita personal income after a number of iterations corresponding to the half-life

Figure A5: Transitional dynamics: 1991:Q1-2000:Q4 (a)



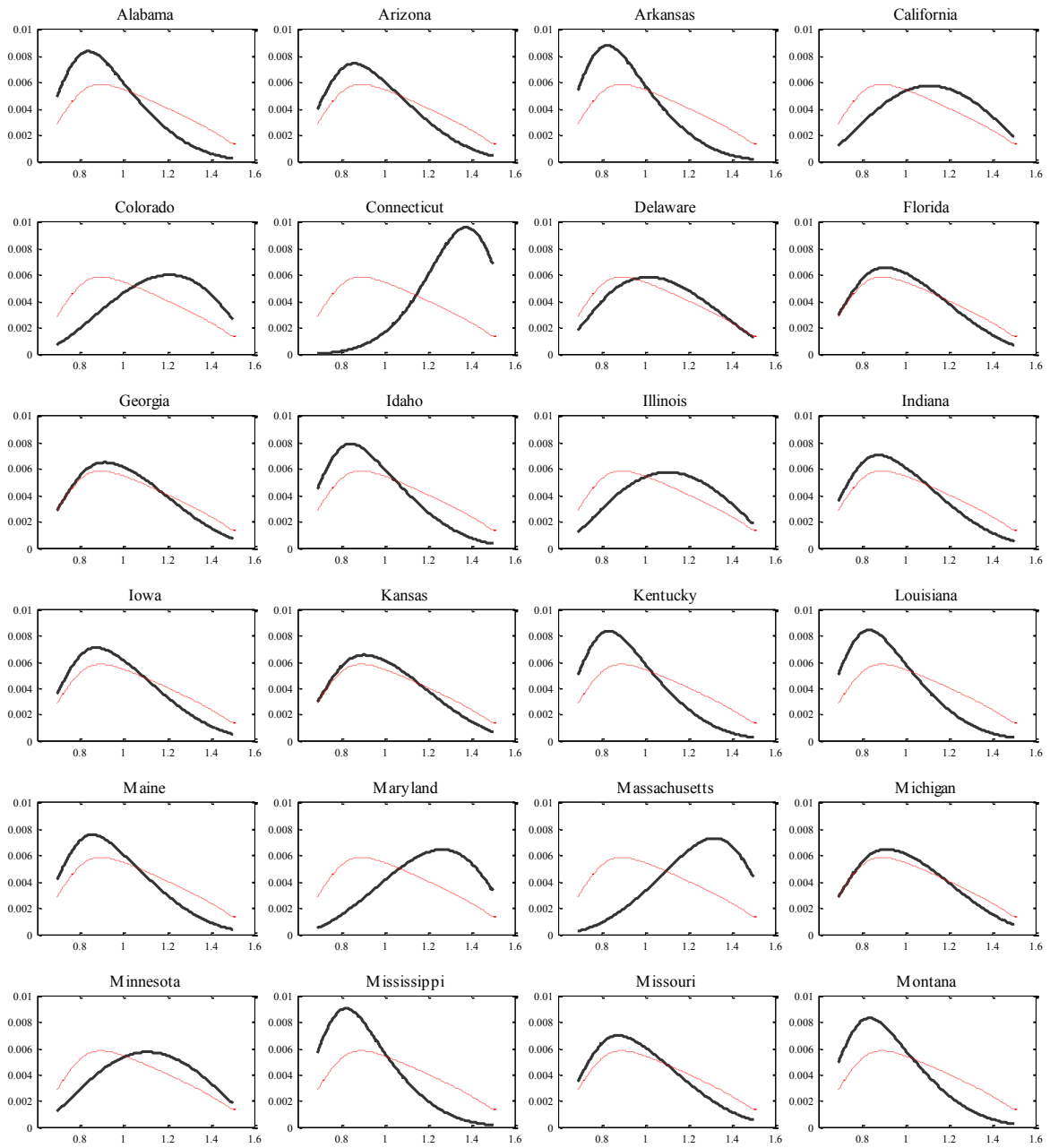
Notes: Plots of the conditional density distributions for each state's initial level of (HP-filtered) per capita personal income after a number of iterations corresponding to the half-life

Figure A6: Transitional dynamics: 1991:Q1-2000:Q4 (b)



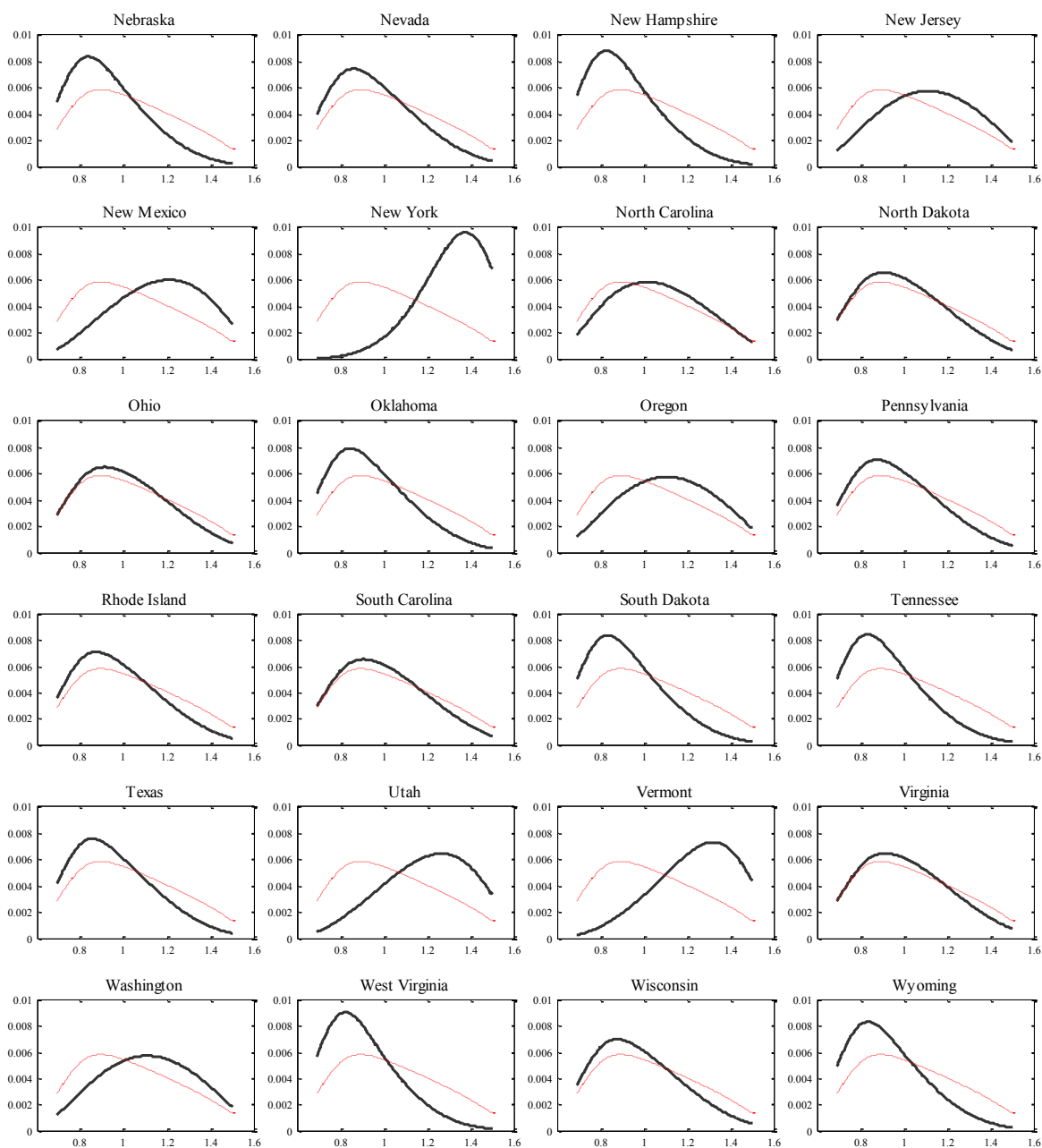
Notes: Plots of the conditional density distributions for each state's initial level of (HP-filtered) per capita personal income after a number of iterations corresponding to the half-life

Figure A7: Transitional dynamics: 2001:Q1-2010:Q4 (a)



Notes: Plots of the conditional density distributions for each state's initial level of (HP-filtered) per capita personal income after a number of iterations corresponding to the half-life

Figure A8: Transitional dynamics: 2001:Q1-2010:Q4 (b)



Notes: Plots of the conditional density distributions for each state's initial level of (HP-filtered) per capita personal income after a number of iterations corresponding to the half-life