Co-funded by the
Erasmus+ Programme
of the European Union

Project funded by
European Commission Erasmus + Programme –Jean Monnet Action
Project number 553280-EPP-1-2015-1-IT-EPPJMO-MODULE

Advanced Spatial Nonparametric Techniques

Margherita Gerolimetto and Stefano Magrini

Department of Economics - Ca'Foscari University of Venice

Pisa 23 October 2017

Motivation and outlines

Motivation

- it is quite common in convergence analyses across spatial units (countries, regions) that data exhibit strong spatial dependence
- neglecting spatial dependence may affect the results

$\Rightarrow$ Study (develop) tools for the analysis of cross-sectional convergence within the distribution dynamics approach when data are spatially dependent

Outline

- what is distribution dynamics (DD)? (refresh)
- What are the consequences of spatial dependence on the analysis of distribution dynamics?
- develop a two-step spatial nonparametric estimator (SNP) for adjusting existing tools in distribution dynamics analysis
  - what is a nonparametric regression?
  - what is SNP and how it is plugged into DD?

Distribution dynamics

The distribution dynamics approach in short

- let $F(Y_t)$ and $F(Y_{t+s})$ represent the cross-sectional distributions of per capita income at time t and $t+s$
- assume they admit a density ($f(Y_t)$ and $f(Y_{t+s})$ respectively)
- assuming the dynamics between time t and $t+s$ can be modelled as a first order process, then

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

- convergence is analysed through:
  - an estimate of the **conditional density** (or stochastic kernel) $f(Y_{t+s}|Y_t)$, traditionally obtained via the **kernel estimator**
  - an estimate of the **ergodic** (or stationary) **distribution** (as $s \to \infty$), under the assumption that the process is Markov and time homogeneous

Conditional density

The corner-stone of the approach is the **conditional density**

$$f\left(Y_{t+s}|Y_t\right)$$

or, using a more general notation,

$$f\left(Y|X\right)$$

a non parametric estimate of $f\left(Y|X\right)$ can be

$$\hat{f}\left(Y|X\right) = \frac{\hat{f}\left(X,Y\right)}{\hat{f}\left(X\right)}$$

this is equivalent to the kernel density estimator

$$\hat{f}(Y|X) = \sum_{j=1}^{n} w_j(X) K_b\left(Y - Y_j\right) \quad with \quad w_j(x) = \frac{K_a\left(X - X_j\right)}{\sum_{j=1}^{n} K_a\left(X - X_j\right)}$$

The mean of the conditional density $f\left(Y|X\right)$ is $E(Y|X)$, the **mean function** $M(X) \rightarrow$ nonparametric regression!

Note: with the alternative notation, the mean of the conditional density $f\left(Y_{t+s}|Y_t\right)$ is the **mean function**, $M(Y_t)$)

Conditional density estimation: *mean-bias* issue

Hyndman *et al.* (1996)

- the mean function estimator implicit in the traditional kernel estimator of the conditional density is the **local constant estimator**
- the bias of the mean function estimate is carried over onto the conditional density estimate (*mean-bias*)
- the local constant estimator has poor bias properties

$\Rightarrow$ the local constant estimator can be replaced with other smoothers employed in nonparametric regressions $Y = M(X) + \epsilon$ (*mean-bias adjustment*)

Conditional density estimation: Hyndman *et al.*'s correction

- Hyndman *et al.*'s proposal is

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

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

- the mean bias of the estimator $\hat{f}^*(Y|X)$ depends on the mean bias of the smoother employed to obtain $\hat{M}(X)$:
  - Nadaraya-Watson smoother $\rightarrow$ traditional kernel regression, stronger bias
  - lower mean bias can be achieved by employing smoothers with better bias properties (e.g. local linear)

Since we are analyzing economic convergence dynamics

$\Rightarrow$ the mean function estimate required in the adjustment procedure is in fact an autoregression $Y_{t+s} = M(Y_t) + \epsilon_t$

$\Rightarrow$ the error terms are likely to be spatially dependent

The spatial dependence issue

Note that

- the statistical properties of $\hat{M}(Y_t)$ assume errors are zero mean and **uncorrelated**
- however, in growth and convergence studies data exhibit **spatial dependence**

$\Rightarrow$ consequences of neglecting spatial dependence in the estimate of $M(Y_t)$ are also carried over onto the conditional density estimate

Within the distribution dynamics framework

- the issue is (only rarely) tackled via spatial filtering
- we opt for an approach that preserves the information brought by spatial dependence, including it into the estimation process
  - per capita income in a US state is correlated to that observed in neighboring states
  - states' mobility within the cross-sectional distribution of per capita income is significantly affected by the position of geographical neighbors within the same distribution (Rey, 2001)

Spatial NonParametric (SNP) regression

SNP is a two-step procedure for nonparametric regression with spatially dependent data whose specific features are:

- it does not require *a priori* parametric assumptions on spatial dependence
- the information on the dependence structure is drawn from a nonparametric estimate of the spatial covariance matrix, called **spline correlogram** $\rightarrow$ W-freee estimate!!!

In addition:

- can be employed to estimate the mean function required in Hyndman's *mean-bias* adjustment, thus providing a way of dealing with both the *mean-bias* and the spatial dependence issues

SNP procedure

Objective: estimate $Y = M(X) + u$

Tool: SNP procedure

0. *Pilot fit*: estimate $M(X)$ with a local polynomial smoother to obtain $\hat{u} = Y - \hat{M}(X)$
1. *Nonparametric covariance matrix estimation*: use the **spline correlogram** to obtain $\hat{V}$, the estimated **spatial covariance** matrix of $\hat{u}$ (using, simply, a distance matrix)
2. *Final fit*: run the **modified regression** $Z = M(X) + \epsilon$ where
   - $Z = \hat{M}(X) + L^{-1}\hat{u}$ replaces Y
   - L is obtained through the Cholevsky decomposition of $\hat{V}$

   $\Rightarrow$ residuals ϵ are free from spatial dependence

Properties

- asymptotic properties are derived by adapting Martins-Filho and Yao's (2009) theoretical framework
  - A consistent estimate of $\hat{V}$ is the requirement for the two-step procedure being sound in terms of asympotical statistical properties
- finite sample properties are established through a Monte Carlo experiment

Spatial correlation function

A spatial covariance/correlation function can be defined as follows:

$$\gamma(s_i, sj) = \sigma^2 f(d_{ij}) \qquad \rho(s_i, sj) = f(d_{ij})$$

where:

- d_{ij} is the distance between sites i,j
- $f(.)$ is a decaying function such that $\frac{\partial f}{\partial d_{ij}} < 0$, $|f(d_{ij}| \leq 1$

To sum up, the full matrix of spatial correlation

- has elements $\rho(s_i, sj) = f(d_{ij})$, where $|\rho(s_i, sj)| \leq 1$ for each i and j
- must be positive semidefinite

Spline correlogram (Bjørnstad and Falk, 2001)

The spline correlogram is a continuous nonparametric positive semidefinite estimator of the covariance function:

- start from the sample correlation

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

- take a cubic B-spline K

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

  The advantage in using the B-spline is in that this smoother adapts better to irregularly spaced data and produces a consistent estimate of the covariance function.

- since $\tilde{\rho}$ must be not only consistent, but also positive semidefinite, use the Fourier-filter

Spline correlogram: more in-depth 1

The smoothing spline solves the fitting problem by selecting the function f that minimizes the penalized residual sum of squares (RSS)

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

1 measures closeness to the data
2 penalizes curvature in the function, τ is a fixed smoothing parameter and 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:

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

where, d is a generic measure of distance and a is a tuning parameter.

Spline correlogram: more in-depth 2

In addition, since the estimator $\tilde{\rho}(s_i, s_j)$ must be not only consistent but also positive semidefinite a Fourier-filter is adopted (Hall *et al.*, 1994).

The latter works as follows:

1. Calculate the Fourier transform of $\tilde{\rho}(s_i, s_j)$
2. all negative excursions of the transformed function are set to zero
3. otbain by backtransformation a nonparametric positive semidefinite estimate of the spatial correlation function

To sum up

- SNP is a tool for nonparametric regression when data are spatially dependent
  - it is based on a nonparametric estimate of the spatial correlation structure, hence leading us to work in a W-free framework
- SNP can be used to estimate the mean function within Hyndman's *mean-bias* adjustment thus improving the properties of the conditional density estimator