Spatial Long Memory

Peter M. Robinson

London School of Economics

Seminar, Universita di Verona, February 13 2019

'Spatial long memory' goes back at least to:

Smith, H. Fairfield (1938). An empirical law, describing heterogeneity in the yields of agricultural crops. Journal of Agricultural Science 28, 1–23.

Consider n agricultural plots on a field.

Yield at location $i = (i_1, i_2)$ is $x_{i_1 i_2}$

Sample mean $= \overline{x}$

Suppose the i, j locations of the data form a regular lattice (equal spacing),

x x x x x x x x x x x x x x x

x x x x x x x x x x x x x x x

x x x x x x x x x x x x x x x

x x x x x x x x x x x x x x x

x x x x x x x x x x x x x x x

x x x x x x x x x x x x x x x

x x x x x x x x x x x x x x x

Smith assumed in effect:

$$Var\,(\overline{x}) \sim Cn^{2d-1},\ 0 \le d < 1/2, \tag{1}$$

and estimated d by LS regression after logging.

If the lattice has dimension $n^{1/2} \times n^{1/2}$, the model (1) is consistent with the underlying isotropic model

$$Cov\left(x_{i_1 i_2}, x_{j_1 j_2}\right) \sim c\sqrt{(i_1 - j_1)^2 + (i_2 - j_2)^2}^{2d-1} = c\,\|i - j\|^{2d-1},$$

which is equivalent to the usual autocovariance function of a stationary long memory time series with differencing parameter d, eg FARIMA(p, d, q).

Since then, many papers on 'spatial long memory', but topic has not been developed as systematically or comprehensively as 'long memory time series '.

Some distinctive issues arising in the 'spatial' case:

Isotropy or not?

Regular or irregular sampling?

Unilateral or multilateral, curse of dimemnsionality?

Edge effect?

Topics:

1. Inference on location and regression with long memory errors

2. Inference on second-order properties of long memory stationary processes

3. Miscellaneous topics: nonstationary processes, irregular spacing, adaptive estimation, nonparametric regression

We do not consider 'spatial autoregressive'-type models (which depend on a user-chosen weight matrix of geographic or economic inverse distances); these typically have short memory.

1 Inference on location and regression with long memory errors

Sample mean $\overline{x}$ is a basic statistic, whose asymptotic properties are well known under time series short and long memory.

In time series case, suppose

$$x_t = \mu + \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} \alpha_j^2 < \infty,$$

where $\{\varepsilon_t\}$ is a sequence of iid (0,1) random variables, or even homoscedastic (but not necessarity id) martingale differences.

Implies x_t is stationary, and can have 'short memory'.or 'long memory' or 'negative memory'.

First suppose x_t has short memory, or is $I(0)$, i.e. x_t has spectral density

$$f(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} Cov\left(x_t, x_{t+j}\right) \cos(j\lambda), \quad -\pi \leq \lambda \leq \pi,$$

that satisfies

$$0 < f(0) < \infty.$$

Then $\overline{x}$ is an asymptotically normal and efficient estimate of μ,

$$\sqrt{n}\left(\overline{x} - \mu\right) \rightarrow_d \mathcal{N}\left(0, 2\pi f(0)\right), \text{ as } n \rightarrow \infty,$$

where $f(0)$ can be consistently estimated, e.g. by nonparametric spectral estimate.

Long and short memory can be described by $I(d)$ process with

$$\alpha_j \sim S(j) j^{d-1}, \; 0 < d < 1/2,$$

for a slowly varying function $S(j)$ and

$$\begin{aligned} 0 &< d < 1/2, \text{ so } f(0) = \infty, \text{ long memory,} \\ -1/2 &< d < 0, \; f(0) = 0, \text{ negative memory.} \end{aligned}$$

There is interest in allowing for a nonconstant $S(j)$ (eg $S(j) = \log j$), and this has been done in limit theory for basic statistics, and also in (semiparametric) estimation of d.

But we focus on constant $S(j)$, which covers FARIMA and FBM.

$\overline{x}$ is no longer asymptotically efficient (Adenstedt, 1975) but it can still be asymptotically normal, albeit possibly with a different rate of convergence.

We can obtain

$$n^{1/2-d}\left(\overline{x}-\mu\right)\rightarrow_{d}\mathcal{N}\left(0,\sigma^{2}\left(d\right)\right)\text{ as }n\rightarrow\infty,\ 0<d<1/2,$$

$$n^{1/2}\left(\overline{x}-\mu\right)\rightarrow_{d}\mathcal{N}\left(0,\sigma^{2}\left(d\right)\right)\text{ as }n\rightarrow\infty,\ -1/2<d<0.$$

We can studentize, with consistent estimates of d and $\sigma^2(d)$ (Robinson, 2005a).

More generally consider the self-normalized statistic

$$u = (\overline{x} - \mu) / \left\{ E \left(n^{-1} \sum_{t=1}^{n} (x_t - \mu) \right)^2 \right\}^{1/2}.$$

Ibragimov and Linnik (1971) showed that

$$u \rightarrow_d \mathcal{N}(0,1), \text{ as } n \rightarrow \infty,$$

if only the ε_t are iid (relaxable) and

$$\sum_{j=0}^{\infty} \alpha_j^2 < \infty, \; E \left(\sum_{t=1}^{n} (x_t - \mu) \right)^2 \rightarrow \infty, \text{ as } n \rightarrow \infty.$$

Note that under negative memory

$$E\left(\sum_{t=1}^{n}(x_t-\mu)\right)^2$$

at best diverges slowly, indeed if $y_t = \mu + \varepsilon_t - \varepsilon_{t-1}$

$$\sum_{t=1}^{n}(x_t-\mu) = \varepsilon_n - \varepsilon_0$$

and it does not diverge at all.

For spatial data, consider an $m-$dimensional lattice.

t is a multiple index $(t_1, ..., t_m)$ with $t_j \in \mathbb{Z} = \{0, \pm 1, ...\}$, $j = 1, ..., m$.

Consider covariance stationary process x_t observed on the lattice

$t \in \mathbb{N} = \{t : 1 \leq t_i \leq n_i,\ i = 1, ..., m\}$, $n = \prod_{i=1}^{m} n_i$

(or more general region):

$$x_t = \sum_{j \in \mathbb{Z}^m} \alpha_{t-j} \varepsilon_j, \quad \sum_{j \in \mathbb{Z}^m} \alpha_j^2 < \infty, \quad \varepsilon_j \sim (0, 1), \text{ inid.}$$

E.g. in isotropic case, $m = 2$

$$\alpha_j \sim \|j\|^{-\delta} = \left(j_1^2 + j_2^2\right)^{-\delta/2},$$

with memory parameter $\delta > 1$.

For time series long memory has been 'explained' as arising from cross-sectional aggregation of AR(1) processes, see Robinson (1978), Granger and (1980) etc.

This interpretation has been extended to spatial processes by eg Lavancier (2011).

Let Λ be the domain of observations and $\lambda = Vol\,(\Lambda)$, so $\lambda \sim n_1 n_2$ in 2-dimensional rectangular lattice case

For

$$S = \sum_{t \in \Lambda} x_t$$

we have

$$\lambda^{\delta-3} S \rightarrow_d \mathcal{N}\left(0, \sigma^2\,(\delta)\right),$$

with

$$\begin{aligned} \delta &\in (1,2) \text{ (long memory)}, \\ \delta &\in (2, 2.5)\,, \sum_{j \in \mathbb{Z}^2} \alpha_j = 0 \text{ (negative memory)}. \end{aligned}$$

Also

$$\lambda^{-1/2} S \rightarrow_d \mathcal{N}\left(0, \sigma^2(\delta)\right),$$

with

$$\delta > 2.5, \quad \sum_{j \in \mathbb{Z}^2} \alpha_{jk} = 0.$$

Studentization?

Extension to regression models:

An interesting case, for $m = 2$, is

$$y_t = \beta_0 + \beta_1 t_1^{\theta_1} + \beta_2 t_2^{\theta_2} + x_t,$$

because regression powers will 'interact' with memory parameters.

For a more general model, and with general m, Robinson (2012a) considered nonlinear LS estimation of β's and $\theta' s$, for short memory x_t.

In this case, estimates are efficient, but β's are estimated slightly less well than θ's, with rates dependng on θ's (and slightly less well than if θ's were known) .

With long memory x_t rates will be slower, and if memory is strong enough relative to magnitude of θ's we may not be able to estimate some β's and $\theta' s$, because error term x_t dominates regression component.

2 Inference on second-order properties of long memory stationary processes

Now consider estimating either a full parametric model for the spectrum/autocovariance function, or else a 'semiparametric' one specified for only low frequencies/long lags.

For short memory time series, early work on asymptotics covered LS and YW estimates of stationary AR.

Continuous and discrete frequency, and Gaussian PMLE, estimates of ARMA and other short memory time series were well covered by Hannan (1973).

He relaxed iid assumptions on innovations, allowing them, and centred squared innovations, to be stationary martingale differences (expressing the natural ordering of time series data), with only finite second moments required (for estimation of dependence parameters), expressing the natural ordering of time series data.

(In fact the squared innovations need only be absolutely integrable).

His CLT prooof was essentially found to work under long memory with differencing parameter $d \in (0, 1/4)$ by Yajima (1985).

Using a different method of proof Fox and Taqqu (1986) established a CLT for $d \in (0, 1/2)$.

For spatial lattice data, a vital early (short memory) reference is Whittle (1954).

Noting that multilateral MA/AR representations are more natural for spatial data than for the unilateral/one-sided ones normal in time series, he pointed out identifiability problems with multilateral repesentations, and extended the (unilateral) Wold representation for purely nondeterministic time series to spatial processes, introducing 'half-plane' representations.

But a given multilateral ARMA doesn't necessarily have a neat half-plane representation, where AR or MA orders may be infinite.

Sometimes a 'quarter plane' representation is possible.

Whittle, and others (eg Martin (1979), Tjostheim (1983), Kashyap (1984), Huang and Anh (1992)) considered estimation of various short memory parametric models, some extending ARMA in an isotropic or separable way, eg, for $m = 2$,

$$(1 - \alpha_1 L_1)(1 - \alpha_2 L_2) x_t = \varepsilon_t,$$

with iid ε_t.

Other kinds of models, eg Matern, were considered by eg Stein (1999).

For asymptotic theory of estimates one basic question is whether to use increasing domain (as usually in time series) or fixed-domain (infill) asymptotics.

Infill asymptotics can lead to results that are not easy to use in practice and even to inconsistency.

Also one can consider a compromise, where domain increases slowly.

Another issue is the 'edge effect'.

Consider estimating the autocovariance $\gamma_j = cov\left(x_t, x_{t+j}\right)$ of a (zero mean) process x_t by

$$\widehat{\gamma}_j = \frac{1}{n} \sum_{t,t+j \in \Lambda} x_t x_{t+j}.$$

Various estimates of parameters (and indeed nonparametric spectral estimates) are essentially functions of the $\widehat{\gamma}_j$.

For $m = 1$ (time series) Λ contains $n - j$ points and so (for fixed j), $\widehat{\gamma}_j$ has bias $O\left(n^{-1}\right)$,so doesn′t prevent CLT for $n^{1/2}\left(\widehat{\gamma}_j - \gamma_j\right)$ from holding.

But for $m = 2$, and $\Lambda = \left[1, n^{1/2}\right] \times \left[1, n^{1/2}\right]$ bias is of exact order $n^{-1/2}$ so limit distribution of $n^{1/2}\left(\widehat{\gamma}_j - \gamma_j\right)$ has nonzero mean.

And for $m = 3$ and $\Lambda = \left[1, n^{1/3}\right] \times \left[1, n^{1/3}\right] \times \left[1, n^{1/3}\right]$ bias is of exact order $n^{-1/3}$ so no CLT.

Likewise for $m > 3$.

So obvious extensions of time series estimates (eg Whittle) of parameters don't work.

Solutions;

1. (Guyon (1982)) Base estimation on unbiased estimates

$$\widetilde{\gamma}_j = n\widehat{\gamma}_j / \# \Lambda$$

2. (Dahlhaus and Kuensch (1987)) Tapering of x_t.

3. (Robinson and Vidal Sanz (2006), Robinson (2007)) Trimming: omitting $\widetilde{\gamma}_j$ for large $\|j\|$.

Edge effect may be an even bigger issue for long memory spatial processes.

Nevertheless a number of parametric long memory stationary linear models, both isotropic and separable, have been considered, with some asymptotic theory.

E.g. Sethuraman and Basawa (1995), Boissy et al (2005), Shitan (2008), Beran et al (2009).

For cyclic/seasonal long memory models see eg Cisse et al (2016).

For 'semiparametric' estimates of the memory parameter:

Log periodogram spatial estimates (extending Geweke and Porter-Hudak (1983)) have been considered by eg Ghodsi and Shitan (2016), the latter extending asymptotic theory of Robinson (1995a) (G and S assume a parametric model).

Local Whittle spatial estimates (extending Kuensch (1986)) have been developed by Guo et al (2009), Durastanti et al (2014), extending asymptotc theory of Robinson (1995b).

Scope for further study of issues of model choice (especially due to danger of 'curse of dimensionality' in lattice models) and bandwidth choice.

3 Miscellaneous topics

3.1 Nonstationary processes

Really, time series AR(1) with a unit root is a 'long memory' process because it has longer memory than a stationary long memory process, and it coincides with a fractional process with $d = 1$.

But nesting it in the fractional, rather than AR, class leads to standard asymptotics, eg CLT for memory and other parameter estimates, even under nonstationarity, using either tapering and skipping of frequencies in discrete-frequency Whittle estimation, or using CSS estimation.

These ideas seem capable of extension to nonstationary fractional processes on a spatial lattice.

Current work on semiparametric estimation for underlying continuous isotropic process for $m = 2$, using tapering and skipping, with Yajima and Matsuda.

We can have multivariate observations, and consider cointegration etc of spatial data.

3.2 Irregular spacing

This is more likely to be an issue with spatial data than with time series data.

For an underlying continuous process, we can in principle construct a Gaussian likelihood, conditional on the observation points.

But conditions for asymptotic theory are messy relative to the regular spaced case.

Sometimes irregular spacing is due to missing from a regular lattice.

Then we might extend Parzen's (1963) amplitude modulating idea for missing data in time series, studying the process $a_t x_t$, where a_t is a zero-one process.

3.3 Adaptive estimation in semiparametric models

'Gaussian' estimates, such as Whittle, of course retain their property of robustness to departures from Gaussianity with spatial data.

But if the process is non-Gaussian such estimates are inefficient.

For time series data 'adaptive' estimation has been developed, where efficiency is established under a nonparametric error distribution, eg for long memory series Hallin and Serroukh (1998), Robinson (2005b).

These ideas seem extendable to long memory lattice processes.

3.4 When ordering doesn't matter

Unlike with time series data there is no natural uni-dimensional ordering of spatial data.

For spatial lattice there is typically an ordering of locations only with respect to each of the m dimensions.

For inference on many features, such as spatial dependence parameters, respect for the ordering is crucial.

But for estimating some 'instantaneous' features, such as location, and stochastic design nonparametric regression, ordering can be disregarded and the spatial data x_i mapped arbitrarily into a sequence u_i, $i = 1, ..., n$.

We might assume, say (Robinson, 2012b),

$$u_i = \sum_{j=1}^{\infty} a_{ij} \eta_j, \quad 1 \le i \le n, \quad \sup_i \sum_{j=1}^{\infty} a_{ij}^2 < \infty. \tag{2}$$

We have 'long memory' if

$$\frac{1}{n} \sum_{i,j=1:i \neq j}^{n} Cov(u_i, u_j) = \frac{1}{n} \sum_{i,j=1:i \neq j}^{n} \sum_{k=1}^{\infty} a_{ik} a_{jk} \to \infty, \text{ as } n \to \infty.$$

For stochastic design nonparametric (kernel) regression, with spatial data, Robinson (2012b) gave conditions for CLT, based partly on (2) and on distance between multivariate and product univariate densities, which can cover long memory.