Some nonparametric regression problems for Gaussian subordinated time series observations with long-memory

Ghosh, Sucharita
Swiss Federal Research Institute WSL, Statistics Lab
Zuercherstrasse 111
8903 Birmensdorf, Switzerland
E-mail: rita.ghosh@wsl.ch

Consider the nonparametric regression problem

$$y_j = g(t_j) + e_j, \; j = 1, 2, ...n \tag{1}$$

where y_j are time series data observed at evenly spaced time points, g is smooth, $t_j = j/n$ are rescaled times and the errors e_j have zero mean and finite variance. Moreover, the errors are assumed to depend on an unobserved stationary Gaussian process Z_i via an arbitrary transformation G. In other words, the following holds (Taqqu 1975):

$$e_j = G(Z_j) \tag{2}$$

We assume that G is an arbitrary Lebesgue-measurable $\mathbb{L}^2$ function with respect to the standard normal density and that it allows for a Hermite polynomial expansion with Hermite rank $m \geq 1$. By definition, the errors e_j are stationary but their marginal distributions depend on the exact form of G. When G is the identity function $G(x) = x$, (1) will have Gaussian errors. Typically however this will not be the case. The form of G will be fairly arbitrary for many data sets. In some applications, the marginal probability distribution of the errors may be time dependent as for instance in the model

$$e_j = G(Z_j, t_j) \tag{3}$$

Nonparametric regression models of these types have been considered by various authors for both discrete and continuous time processes. Examples include estimation of the exceedance probabilities and marginal distributional quantiles as functions of time, or for estimating the points of time when rapid changes occur in the trend curve. See for instance Draghicescu (2002) and Menendez (2009) and references therein.

This paper focuses on time series replicates that have a common trend and the errors are Gaussian subordinated. This has applications for instance in estimating overall trend in instrumental climate records from several locations or in finding certain common temporal features in deep core data that occur in palaeo environmental sciences. In this paper we consider independent replicates although generalizations are possible. Consider the nonparametric regression model defined for discrete time processes

$$y_j^{(i)} = g(t_j) + e_j^{(i)}, \; i = 1, 2, ..., k, \; j = 1, 2, ..., n \tag{4}$$

where i denotes replication number and j is a point of time. The errors $e_j^{(i)}$ are Gaussian subordinated as,

$$e_j^{(i)} = G(Z_j^{(i)}, i) = \sum_{l=m}^{\infty} \frac{c_l^{(i)}}{l!} H_l(Z_j^{(i)}) \tag{5}$$

In (5), $c_l^{(i)}$ are Hermite coefficients and satisfy some mild regularity conditions. The Hermite rank m is a positive integer for which $c_m^{(i)}$ is non-zero and $c_l^{(i)} = 0, \; l < m$. In general, m may vary between replicates. For simplicity, we let the Hermite rank be the same for all replicates, so that the Hermite polynomial expansions for the regression errors in the different replications have the same leading terms, except possibly for differing values of the leading Hermite coefficients. The errors will be assumed to be independent between replicates. As for the latent Gaussian processes, $\{Z_j^{(i)},\ i = 1, 2, ..., k\}$ are $k \geq 1$ stationary zero mean Gaussian processes with

$$\begin{aligned} cov\left(Z_j^{(i)}, Z_{j'}^{(i')}\right) &= 0,\ i \neq i' \\ &= \gamma_i\left(|j - j'|\right),\ i = i' \end{aligned}$$

where $\gamma_i\left(|j - j'|\right)$ is characterized via a fractional differencing parameter $\delta_i \in (-1/2, 1/2)$ and a positive constant C_i as follows:

- $f_i(\lambda) \sim C_i|\lambda|^{-2\delta_i}$ as $\lambda \to \infty$ is the spectral density of $Z^{(i)}$ and
- The spectral density above implies the covariances $\gamma_i(u) \sim D_i|u|^{2\delta_i - 1}$, $\delta_i \neq 0$ where,

$$D_i = C_i sin(\pi\delta_i)\Gamma\left(1 - 2\delta_i\right)/(1 + 2\delta_i).$$

In particular, $\delta_i > 0$ implies long-range dependence in Z_i in which case, $\sum_{u=-\infty}^{\infty} \gamma_i(u) = \infty$ and $f_i(\lambda)$ has a pole at zero. Our aim lies in estimating the trend g and its derivatives by smoothing the sample mean

$$\bar{y}_j = \frac{1}{k}\sum_{i=1}^{k} y_{i,j}$$

using for instance the kernel estimator

$$\hat{g}^{(\nu)}(t) = \frac{1}{nb}\sum_{j=1}^{n} K^{(\nu)}\left(\frac{t_j - t}{b}\right)\bar{y}_j \qquad (6)$$

for $\nu \geq 0$ and $K^{(\nu)}$ is a suitably defined kernel (e.. Gasser & Mueller 1984). More generally, $\bar{y}_j$ in (6) may be replaced by a locally smoothed estimator. For instance if the replicates result from the use of a covariate as for instance in a randomized block design, values of the covariate can be used for estimation of $\bar{y}_j$ for a replicate group. The focus here is investigation of the role of the number of replicates k in relation to the sample size n. Of special interest is the case when the fractional differencing parameters $\delta_1, \delta_2, ..., \delta_k$ are random and have a common distribution function F. It is of interest to know in what way the moments of F play a role. Note that the mean spectral density at the origin can be directly related to the moment generating function of F; see Ghosh (2001).

REFERENCES (RÉFERENCES)

Beran, J. (1994) Statistics for Long-Memory Processes. Chapman & Hall, New York.

Draghicescu, D. (2002) Nonparametric quantile estimation for dependent data. Unpublished Ph.D. thesis no. 2592, EPFL.

Gasser, T., Müller, H.G. (1984) Estimating regression functions and their derivatives by the kernel method. Scandinavian Journal of Statistics, 11, 171-185.

Ghosh, S. (2001) Nonparametric trend estimation in replicated time series. Journal of Statistical Planning and Inference, 97, 263274

Menendez, P. (2009) Statistical Tools for Palaeo Data. Unpublished Ph.D. thesis. Department of Mathematics. Swiss Federal Institute of Technology, Zürich.

Taqqu, M.S. (1975) Weak convergence to fractional Brownian motion and to the Rosenblatt process. Probability Theory and Related Fields, 31, 287-302.