First publ. in: Journal of Multivariate Analysis 102 (2010), 2, pp. 315-335
doi:10.1016/j.jmva.2010.09.007

On spline regression under Gaussian subordination with long memory

Jan Beran *, Arno Weiershäuser

Department of Mathematics and Statistics, University of Konstanz, Germany

A B S T R A C T

Motivated by an example from neurobiology, we consider estimation in a spline regression model with long-range dependent errors that are generated by Gaussian subordination. Consistency and the asymptotic distribution are derived for general Hermite ranks. Simulations illustrate the asymptotic results and finite sample properties. The method is applied to optical measurements of calcium concentration in the antennal lobe of honey bees used in the study of olfactory patterns.

AMS subject classifications:
62M10
62F12
62J02
62P10
62M09

Keywords:
Long-range dependence
Hermite rank
Gaussian subordination
Spline regression
Two-phase regression
Change point

1. Introduction

We consider estimation in a spline regression model with strongly dependent errors. The problem of fitting piecewise polynomials when residuals exhibit long memory is motivated by an application in neurobiology: Galán et al. [26] analyzed the temporal evolution of glomerular activity patterns in the antennal lobe of honey bees (also see [38,27]) which is a central part in the odor processing of insects. Coding of odors is believed to be related to spatiotemporal patterns represented as trajectories (in time) in a multidimensional space, where each dimension accounts for the activity of one glomerulus. Fig. 1(a) shows the activity of a glomerulus of a honey bee, starting 0.2 s after an olfactory stimulus and measured for 24 s at time intervals of 0.04 s. The response to the stimulus fades out fairly quickly. Afterward the process continues to fluctuate randomly at a lower lever, but possibly with a slowly changing mean. The transition between the two phases appears to be (and is expected to be) smooth. Splines provide a natural way of modeling this type of response function. The estimated curve in Fig. 1(a) was obtained by quadratic spline regression with one unknown knot. The estimated knot can be interpreted as the approximate time when the effect of the stimulus ceases to be observable. The distribution of the residuals is close to normal (Fig. 1(c)), but the periodogram exhibits a negative slope in log–log-coordinates which is an indication for long-range dependence (Fig. 1(d)). The question therefore arises which effect long-range dependence has on parameter estimation in spline regression. In this paper the asymptotic distribution of parameter estimates will be derived for error processes obtained by Gaussian subordination.

The occurrence of long memory and related fractal processes in neuroscience is well documented (see e.g. [45]). For other applications and statistical inference for long memory processes; see e.g. [1,23]. A classical reference to splines in general is for instance [20], for splines in statistics; see e.g. [53,24,32]. Diggle and Hutchinson [21] extend Wahba's [52] formulation

* Corresponding author.
E-mail address: jan.beran@uni-konstanz.de (J. Beran).

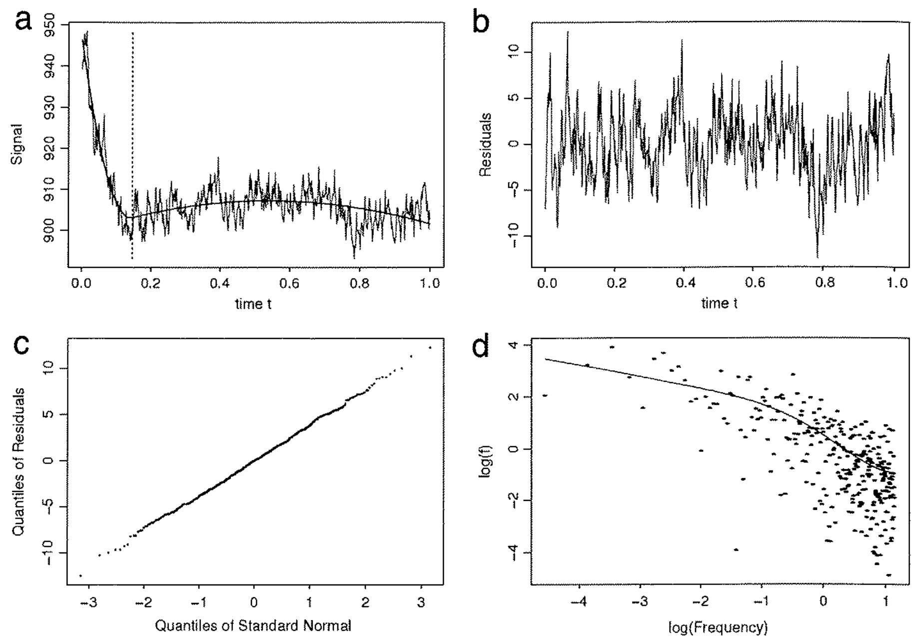

Fig. 1. Optical measurements of the calcium concentration in a glomerulus of the antennal lobe of a honey bee [26] starting 0.2 s after an olfactory stimulus. The fitted trend function was obtained by quadratic spline regression with one free knot η. Also given are the regression residuals, a normal probability plot of the residuals and a log–log-plot of their periodogram together with a fitted spectral density. The dotted vertical line indicates the location of $\hat{\eta}$.

of spline smoothing to correlated errors (also see [40,54]). Multi-phase regression is discussed for instance in monographs by Csörgö and Horvath [16], Brodsky and Darkhovsky [14], Chen and Gupta [15] and Seber and Wild [49]. In the case of i.i.d. observations the least squares estimator is asymptotically normally distributed [29,25]. For multivariate models; see [44,39]. If the knots are known, then the problem of fitting a spline boils down to linear regression. Results on the efficiency of least squares estimators in linear regression with strongly dependent residuals are given in Yajima [55,56], Künsch et al. [43] and Dahlhaus [19] among others. For a review on nonparametric regression with correlated errors; see [47]. Kernel estimators in the long memory context are considered, e.g. in [33,48,17,18]. For local polynomial estimation; see [4–8]. Beran and Ghosh [9] and Gao and Anh [30] consider semiparametric regression models. Koul [41] and Koul and Baillie [42] extend results on M-estimation for long memory processes given in [2] to linear and nonlinear regression with random explanatory variables. Ivanov and Leonenko [35,36] derive asymptotic properties of least squares estimators in continuous time nonlinear regression, Ivanov and Leonenko [37] discuss estimation of the long memory parameter in a similar context.

The results in [35,36] are not directly applicable to our situation, because our parameter space is unbounded and some of the regularity assumptions do not hold. It turns out however that the specific features of spline regression can be used to derive similar results. The paper is organized as follows: In Section 2, basic concepts are introduced briefly, the model and the least squares estimator are defined, and some elementary properties are mentioned. Consistency of the least squares estimator and its asymptotic distribution are derived in Section 3. It is shown that the asymptotic limit can be represented as a weighted average of a Hermite process of an appropriate order. In Section 4, the results are illustrated by simulations and the data example introduced above. Detailed proofs are given in the Appendix.

2. Preliminaries and basic definitions

2.1. The model

We consider the model

$$X(t) = \mu\left(\frac{t}{n}\right) + \xi_t \quad (t = 1, \ldots, n) \tag{1}$$

with $\mu(s)$ denoting a spline function of degree $p \geq 2$. For simplicity of presentation, theoretical results will be given for the case of one unknown knot η. A generalization to multiple knots is straightforward.

A spline function of order p and one free knot has the representation

$$\mu(s) = \sum_{i=1}^{p+2} a_i f_i(s) \quad (s \in [0, 1]) \tag{2}$$

where $a = (a_1, \ldots, a_{p+2})'$ denotes unknown regression coefficients and the functions $f_1, \ldots, f_{p+2}$ are spline basis functions defined as

$$f_1(s) = 1, \qquad f_2(s) = s, \ldots, f_{p+1}(s) = s^p, \qquad f_{p+2}(s) = f_{p+2}(s;\eta) = (s-\eta)_+^p. \tag{3}$$

The error process is assumed to be given by $\xi_t = G(\varepsilon_t)$ where ε_t is a stationary Gaussian process with $\mathbb{E}(\varepsilon_t) = 0$, $\mathrm{Var}(\varepsilon_t) = 1$ and such that

$$\gamma(t-s) = \mathrm{Cov}(\varepsilon_t, \varepsilon_s) = \frac{L(|t-s|)}{|t-s|^{\alpha}} \tag{4}$$

where L is slowly varying at infinity [12] and $\alpha \in (0, 1)$ denotes the long memory parameter. The function $G : \mathbb{R} \to \mathbb{R}$ is a Borel mapping such that $\mathbb{E}[G(\varepsilon_t)] = 0$ and $\mathbb{E}[G^4(\varepsilon_t)] < \infty$. It then follows that G and G^2 have an orthogonal representation in terms of Hermite polynomials (cf. [50,22,46]) given by

$$G(\varepsilon_t) = \sum_{k=m}^{\infty} \frac{C_k}{k!} H_k(\varepsilon_t)$$

and

$$G^2(\varepsilon_t) - \mathbb{E}[G^2(\varepsilon_0)] = \sum_{k=j}^{\infty} \frac{\tilde{C}_k}{k!} H_k(\varepsilon_t)$$

where

$$H_k(x) = (-1)^k \exp\left(\frac{1}{2}x^2\right) \frac{\mathrm{d}^k}{\mathrm{d}x^k}\left[\exp\left(-\frac{x^2}{2}\right)\right],$$

and $m, j \geq 1$ denote the Hermite ranks of G and $G^2 - \mathbb{E}(G^2)$, respectively. The covariance structure of ξ_t is given by

$$\mathrm{Cov}(\xi_t, \xi_s) = \sum_{k=m}^{\infty} \frac{C_k^2}{k!} \gamma^k(s-t). \tag{5}$$

In particular, assuming $m\alpha < 1$ and $j\alpha < 1$ we have

$$\mathrm{Var}\left(n^{-1}\sum_{t=1}^{n} G(\varepsilon_t)\right) = O(\gamma^m(n)), \qquad \mathrm{Var}\left(n^{-1}\sum_{t=1}^{n} G^2(\varepsilon_t)\right) = O(\gamma^j(n)).$$

Remark 1. In order to derive the asymptotic distribution, we essentially only need the condition $\mathrm{Var}\left(n^{-1}\sum_{t=1}^{n} G^2(\varepsilon_t)\right) \to 0$ which still holds if $j\alpha \geq 1$. Therefore, all the basic results remain the same, only the rates of convergence in some intermediate calculations change. In contrast, the results change significantly if $m\alpha > 1$. In this case $\mathrm{Var}\left(n^{-1}\sum_{t=1}^{n} G(\varepsilon_t)\right) = O(n^{-1})$ and $\sum_{t=1}^{n} G(\varepsilon_t)$ is asymptotically Gaussian. The limiting behavior is then in the realm of results by Breuer and Major [13].

2.2. Least squares estimation

The parameter vector to be estimated is $\theta = (a', \eta)' \in \Theta = \mathbb{R}^{p+2} \times (0, 1)$. Note that $\eta \neq 0$ and 1 in order to keep the model identifiable. A further identifiability condition is $a_{p+2} \neq 0$. The least squares estimator of θ is defined by

$$L(\hat{\theta}) = \inf_{\tau \in \Theta} L(\tau) \tag{6}$$

where

$$L(\tau) = \sum_{t=1}^{n}\left[X(t) - \mu\left(\frac{t}{n}, \tau\right)\right]^2.$$

The following definitions will be needed. For given η, set $w_{i,j} = f_j\left(\frac{i}{n}\right)$ $(1 \leq i \leq n;\ 1 \leq j \leq p+2)$ and define the $n \times (p+2)$ matrix

$$\mathbf{W_n} = \mathbf{W_n}(\eta) = (w_{ij})_{i=1,\ldots,n;j=1,\ldots,p+2}.$$

The columns of $\mathbf{W_n}$ will be denoted by $\mathbf{w_{j,n}}$ $(j = 1, \ldots, p+2)$. For each η, there exists an $N(\eta) \in \mathbb{N}$ such that $\mathbf{W_n'W_n}$ is invertible for all $n \geq N(\eta)$. The projection matrix on the column space of $\mathbf{W_n}(\eta)$ may then be written as

$$P_{\mathbf{W_n}} = P_{\mathbf{W_n}(\eta)} = \mathbf{W_n}(\mathbf{W_n'W_n})^{-1}\mathbf{W_n'}.$$

For parsimonious notation, we set $\mu_n(\theta) = [\mu(t/n, \theta)]_{t=1,\ldots,n}$, $\mathbf{e}_n = (\xi_1, \ldots, \xi_n)$,

$$\Phi_n(\theta_1, \theta_2) = \|\mu_n(\theta_1) - \mu_n(\theta_2)\|_{\mathbb{R}^n}^2 = \sum_{t=1}^{n}\left[\mu\left(t/n, \theta_1\right) - \mu\left(t/n, \theta_2\right)\right]^2,$$

and

$$\Phi_\infty(\theta_1, \theta_2) = \int_0^1 [\mu(u, \theta_1) - \mu(u, \theta_2)]^2 \, du.$$

Note that $\lim_n n^{-1}\Phi_n(\tau_1, \tau_2) = \Phi_\infty(\tau_1, \tau_2)$. Given a vector of observations $\mathbf{X} = [X(1), \ldots, X(n)]'$, the least squares estimate of η is obtained by minimizing $\left\| \mathbf{X} - P_{\mathbf{W_n}(\tilde{\eta})}(\mathbf{X}) \right\|^2$ with respect to $\tilde{\eta}$. Given $\hat{\eta}$, the estimate of $a = (a_1, \ldots, a_{p+2})'$ is obtained by linear regression with design matrix $\mathbf{W_n}(\hat{\eta})$. The estimated mean function is then equal to

$$P_{\mathbf{W_n}(\hat{\eta})}(\mathbf{X}) = P_{\mathbf{W_n}(\hat{\eta})} \left[\mu_n(\theta) + \mathbf{e}_n\right].$$

Since $\eta = 0$ and 1 are excluded from Θ, and for $\hat{\eta} \in \left[0, \frac{1}{n}\right] \cup \{1\}$ the column space of $\mathbf{W_n}(\hat{\eta})$ coincides with the space spanned by polynomials of degree p, we only need to consider $\hat{\eta} \in (1/n, 1)$. Note also that the mapping $\tilde{\eta} \longmapsto P_{\mathbf{W_n}(\tilde{\eta})}\mathbf{X}$ is continuous and constant on the intervals $\left(\frac{1}{n}, \frac{2}{n}\right)$ and $\left(\frac{n-1}{n}, 1\right)$ so that the least squares estimate of η (and a) exists, though it may not be unique.

3. Limit theorems

3.1. Consistency

The following identifiability result is proved in the Appendix.

Lemma 1. *Let $X(t)$ be given by (1), (2) and (4) with true parameter vector $\theta = (a_1, \ldots, a_{p+2}, \eta)' \in \Theta$ such that $a_{p+2} \neq 0$. For an arbitrary $\Delta > 0$ define the set $K(\Delta)$ by*

$$K(\Delta) = \left\{u \in \mathbb{R}^{p+3} : \theta + u \in \bar{\Theta}, \|u\| \geq \Delta\right\}.$$

Then there exists a $\delta > 0$ such that

$$\liminf_n \left(\inf_{u \in K(\Delta)} n^{-1}\Phi_n(\theta, \theta + u) \right) > \delta.$$

We now can prove consistency:

Theorem 1. *Let m, j be the Hermite ranks of G and $G^2 - \mathbb{E}\left[G^2\right]$, respectively. Suppose that $m\alpha < 1$ and $j\alpha < 1$, and define $\kappa = \min(j, m)$. Then, under the assumptions of Lemma 1 we have for any $\Delta > 0$,*

$$\mathbb{P}\left(\left\|\hat{\theta} - \theta\right\| \geq \Delta\right) = O(\gamma^\kappa(n)).$$

3.2. Asymptotic distribution

In the following partial derivatives of μ with respect to θ_j will be denoted by $\mu_{(j)}$. Extending the matrix $\mathbf{W_n}$ to

$$\mathbf{M_n} = \left[\mu_{(j)}(t/n)\right]_{t=1,\ldots,n; j=1,\ldots,p+3} \in \mathbb{R}^{n\times(p+3)}$$

we have

$$\lim_{n\to\infty} n^{-1}(\mathbf{M_n'M_n})_{jk} = \int_0^1 \mu_{(j)}(s, \theta)\mu_{(k)}(s, \theta)ds. \tag{7}$$

In analogy to $\mathbf{W_n}'\mathbf{W_n}$, the matrix $\mathbf{M_n'M_n}$ has full rank if n is large enough so that

$$\Lambda = \lim_n n(\mathbf{M_n'M_n})^{-1}$$

is well defined.

In a first step, $\gamma^{-\frac{m}{2}}(n)(\hat{\theta} - \theta)$ is shown to be asymptotically equivalent to $(\mathbf{M_n'M_n})^{-1}\mathbf{M_n}\mathbf{e}_n$.

Theorem 2. *Under the assumptions of Theorem 1 we have for any $\Delta > 0$,*

$$\mathbb{P}\left(\gamma^{-\frac{m}{2}}(n) \left\|\hat{\theta} - \theta - (\mathbf{M_n'M_n})^{-1}\mathbf{M_n}\mathbf{e}_n\right\| > \Delta\right) = o(1).$$

The asymptotic distribution of $\hat{\theta}$ depends on the first term in the Hermite expansion of G only.

Theorem 3. *Under the assumptions of Theorem 1 we have*

$$\lim_{n\to\infty} \gamma^{m}(n)\mathbb{E}\left[\left(Y_n - (\mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n})^{-1}\mathbf{M}_\mathbf{n}'\mathbf{e}_n\right)^2\right] = 0 \tag{8}$$

and

$$\lim_{n\to\infty} \gamma^{-m}(n)\mathrm{Cov}\left[(\mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n})^{-1}\mathbf{M}_\mathbf{n}'\mathbf{e}_n\right] = \Lambda\Sigma_0\Lambda \tag{9}$$

where Λ is defined above, $C_m = \mathbb{E}[G(\varepsilon)H_m(\varepsilon)]$ is the mth Hermite coefficient of G,

$$Y_n = (\mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n})^{-1}\mathbf{M}_\mathbf{n}'\frac{C_m}{m!}H_m(\varepsilon),$$
$$H_m(\varepsilon) = (H_m(\varepsilon_1), \ldots, H_m(\varepsilon_n))'$$

and

$$\Sigma_0 = \frac{C_m^2}{m!}\left(\int_0^1 \frac{\mu_{(j)}(t)\mu_{(k)}(s)}{|s-t|^{\alpha m}}\mathrm{d}t\mathrm{d}s\right)_{j,k=1,\ldots,p+3}.$$

An immediate consequence of Theorem 3 is that the asymptotic distribution of $\hat{\theta}$ is Gaussian if and only if $m = 1$, whereas for $m > 1$, the asymptotic distribution can be represented by a moving average of a Hermite process.

Corollary 1. *Under the assumptions of Theorem 1 and $m = 1$, we have*

$$\gamma^{-\frac{1}{2}}(n)(\hat{\theta} - \theta) \to_d Z$$

where $Z \sim N(0, \Lambda\Sigma_0\Lambda)$.

Corollary 2. *Under the assumptions of Theorem 1 and $m \geq 2$ we have*

$$\gamma^{-\frac{m}{2}}(n)(\hat{\theta} - \theta) \to_d Z_m$$

where Z_m is a non-Gaussian random variable defined by

$$Z_m = \Lambda\begin{pmatrix} -\int_0^1 \frac{C(m)}{m!}\mathcal{H}_m(t)\,\mathrm{d}\mu_{(1)}(t) + \frac{C(m)}{m!}\mathcal{H}_m(1)\mu_{(1)}(1) \\ \vdots \\ -\int_0^1 \frac{C(m)}{m!}\mathcal{H}_m(t)\,\mathrm{d}\mu_{(p+3)}(t) + \frac{C(m)}{m!}\mathcal{H}_m(1)\mu_{(p+3)}(1) \end{pmatrix}$$

with $\mathcal{H}_m(t)$ denoting a Hermite process of order m.

Corollaries 1 and 2 provide a complete, though complicated, description of the asymptotic distribution. Using a computer algebra system it is possible to obtain closed form formulas for the entries of the covariance matrix $\Lambda\Sigma_0\Lambda$. Consider, for instance, the variance of $\hat{\eta}$: Denote by $P_{\mathbf{W_n}^{\perp}(\tilde{\eta})}$ the projection matrix on the subspace that is orthogonal to the columns of $\mathbf{W_n}^{\perp}(\tilde{\eta})$. Since minimization of

$$\left\|\mathbf{X} - P_{\mathbf{W_n}(\tilde{\eta})}(\mathbf{X})\right\|^2 = \left\|P_{\mathbf{W_n}^{\perp}(\tilde{\eta})}\left(\mathbf{e}_n + \sum_{j=1}^{p+1} a_j\mathbf{w_{j,n}}(\eta) + a_{p+2}\mathbf{w_{p+2,n}}(\eta)\right)\right\|^2$$

with respect to $\tilde{\eta}$ is equivalent to minimizing

$$\left\|P_{\mathbf{W_n}^{\perp}(\tilde{\eta})}\left(\frac{\mathbf{e}_n}{a_{p+2}} + \mathbf{w_{p+2,n}}(\eta)\right)\right\|^2,$$

$\hat{\eta}$ depends on the parameters $(a_1, \ldots, a_{p+2}, \sigma^2)$ via the ratio σ/a_{p+2} only. From the formula for Σ_0 we see that multiplying ξ by a constant c changes the asymptotic variance of $\hat{\eta}$ by the factor c^2. Hence, the asymptotic variance is of the form

$$\mathrm{Var}_{\mathrm{asym}}(\hat{\eta}) = \frac{C_m^2}{m!\,a_{p+2}^2}f(\alpha, \eta)$$

where f is a nonlinear function of η and α.

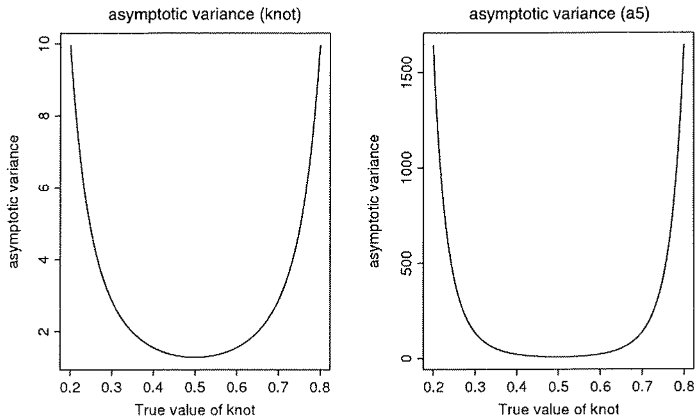

Fig. 2. Asymptotic variance of $\hat{\eta}$ and $\hat{a}_5$ for $\eta = 0.5$, $a_5 = 1$ and $C_1 = 0.01$.

We used Maple 9.5 to derive closed form formulas for $f(\alpha, \eta)$ and $g(\alpha, \eta)$ with $p = 3$. From these formulas it can be seen that for a fixed value of α, the mapping $\eta \mapsto f(\alpha, \eta)$ is symmetric around $\eta = 0.5$. Moreover, f has a pole at $\eta = 0$ of the form

$$f(\alpha, \eta) \sim \frac{1}{\eta^{4+\alpha}} \frac{1200(6 - 2\alpha^2 + 3\alpha)}{(8 - \alpha)(6 - \alpha)(1 - \alpha)_4}$$

with $(1 - \alpha)_k = (k - \alpha)(k - 1 - \alpha) \cdots (1 - \alpha)$. Similarly, the asymptotic variance of $\hat{a}_5$ is of the form $C_m^2/m!g(\alpha, \eta)$, and is thus independent of a. Again, g is symmetric around $\eta = 0.5$ and has poles at the boundary of the form

$$g(\alpha, \eta) \sim \frac{1}{\eta^{6+\alpha}} \frac{7056(8 - 5\alpha^2 + 12\alpha)}{(8 - \alpha)(6 - \alpha)(1 - \alpha)_4}$$

for $\eta \to 0$. Fig. 2 shows plots of g and f with $m = 1$, $\alpha = 0.5$, $C_1 = 0.01$ and $a_5 = 1$, as a function of η. The detailed closed form formulas and the Maple code can be obtained from the authors upon request.

Remark 2. For cubic and higher order splines an explicit rate can be obtained in Theorem 2, as in [36], namely

$$\mathbb{P}\left(\gamma^{-\frac{m}{2}}(n) \left\|\hat{\theta} - \theta - (\mathbf{M}_n'\mathbf{M}_n)^{-1}\mathbf{M}_n\mathbf{e}_n\right\| > \Delta\right) = O\left(\gamma^{\kappa^*}(n)\right),$$

with $\kappa^* = \min\left(j, \frac{m}{2}\right)$. The proof of this sharper result requires a second order approximation that is not applicable to quadratic splines ($p = 2$).

4. Simulations

In this section, the asymptotic results are illustrated by a simulation study. Two different mean functions μ_A and μ_B with $a' = (40, 15, 50, 25, -1500)$ and $\eta = 0.5$ (for μ_A) and $a' = (40, 15, 50, 25, -6000)$ and $\eta = 0.8$ (for μ_B) respectively considered. The error process ξ_i is generated by fractional Gaussian noise with long-range dependence parameter $\alpha \in \{0.4, 0.8\}$ and variance $\sigma^2 \in \{16, 64\}$. Typical sample paths are shown in Figs. 3 and 4.

To see how fast the standardized mean squared error converges to the asymptotic value, we start with a sample size of $n_1 = 6000$ and reduce the number of observations in steps of 5%, i.e. $n_{k+1} = [0.95n_k]$ with $[x]$ denoting the integer part of x. The smallest sample size is $n_{81} = 99$. For each n, 1300 replications are simulated. Figs. 5 and 6 show log–log-plots of the variance of $\hat{\eta}$ and $\hat{a}_5$, respectively. The dashed line corresponds to the normalized asymptotic variance $\gamma(N)\mathrm{Var}_{\mathrm{asym}}(\hat{\eta})$ and $\gamma(N)\mathrm{Var}_{\mathrm{asym}}(\hat{a}_5)$, respectively. Numerical results are given in Tables 1 and 2. Overall, the simulations illustrate consistency of the estimates and convergence of the standardized variance of $\hat{\eta}$ and $\hat{a}_5$ to the asymptotic value derived in Theorem 3. The results also show that convergence is slower for strong long memory ($\alpha = 0.4$) and higher variance ($\sigma^2 = 64$). As predicted by Theorem 3 and the explicit formulas for the asymptotic variance (see Fig. 2), the simulated variances of $\hat{\eta}$ and $\hat{a}_5$ are much smaller when the knot η is placed in the middle ($\eta = 0.5$) (Figs. 7 and 8).

5. Data example

We consider the series introduced in Section 1. The data were provided to us by Giovanni Galizia and Martin Strauch (Department of Biology, University of Konstanz) and are part of a long term project on olfactory coding in insects (see e.g. [38,27,26]). The observations consist of optical measurements of calcium concentration in the antennal lobe of a honey

Table 1
Simulated mean, variance and skewness of $\hat{\eta}$.

n	Mean	Variance V_{sim}	$V_{sim}/\gamma(N)$	$V_{sim}/[\gamma(N)V_{asym}]$	Skewness	Mean	Variance V_{sim}	$V_{sim}/\gamma(N)$	$V_{sim}/[\gamma(N)V_{asym}]$	Skewness
	$\eta = 0.5, \alpha = 0.8, \sigma = 4, a_5 = -1500$					$\eta = 0.5, \alpha = 0.8, \sigma = 8, a_5 = -1500$				
99	0.499	0.002303	0.752	1.036	0.155	0.498	0.011746	3.834	1.320	−0.029
512	0.502	0.000569	0.697	0.960	−0.102	0.502	0.002527	3.092	1.065	0.217
997	0.500	0.000345	0.719	0.991	0.037	0.499	0.001424	2.972	1.023	−0.113
2509	0.500	0.000171	0.746	1.027	−0.028	0.500	0.000619	2.704	0.931	−0.018
3080	0.500	0.000153	0.787	1.085	0.009	0.500	0.000615	3.164	1.090	0.021
4190	0.500	0.000111	0.732	1.008	0.002	0.500	0.000424	2.790	0.961	−0.003
5144	0.500	9.786e−05	0.759	1.046	0.005	0.499	0.000360	2.790	0.961	−0.056
6000	0.500	8.285e−05	0.727	1.002	−0.021	0.501	0.000354	3.109	1.071	−0.001
	$\eta = 0.5, \alpha = 0.4, \sigma = 4, a_5 = -1500$					$\eta = 0.5, \alpha = 0.4, \sigma = 8, a_5 = -1500$				
99	0.497	0.003845	0.0501	1.024	0.011	0.498	0.01737	0.226	1.156	0.015
512	0.501	0.002077	0.0524	1.071	−0.013	0.496	0.00912	0.230	1.175	0.035
997	0.500	0.001511	0.0498	1.017	0.020	0.499	0.00791	0.261	1.331	−0.093
2509	0.500	0.001025	0.0489	0.998	0.026	0.500	0.00467	0.223	1.138	0.118
3080	0.499	0.000918	0.0475	0.971	0.015	0.501	0.00409	0.212	1.081	0.110
4190	0.501	0.000861	0.0504	1.030	−0.062	0.496	0.00381	0.223	1.140	−0.095
5144	0.500	0.000764	0.0486	0.992	−0.102	0.499	0.00347	0.220	1.125	−0.033
6000	0.500	0.000798	0.0539	1.101	−0.040	0.501	0.00316	0.214	1.090	−0.174
	$\eta = 0.8, \alpha = 0.8, \sigma = 4, a_5 = -6000$					$\eta = 0.8, \alpha = 0.8, \sigma = 8, a_5 = -6000$				
99	0.797	0.001325	0.433	1.113	−0.795	0.785	0.009858	3.218	2.070	−2.063
512	0.800	0.000303	0.371	0.954	−0.113	0.796	0.001503	1.839	1.183	−0.817
997	0.800	0.000196	0.408	1.051	−0.216	0.798	0.000794	1.657	1.066	−0.377
2509	0.800	8.581e−05	0.375	0.964	−0.079	0.800	0.000367	1.601	1.030	−0.313
3080	0.800	8.020e−05	0.413	1.062	−0.178	0.799	0.000314	1.616	1.040	−0.230
4190	0.800	6.026e−05	0.397	1.021	−0.179	0.800	0.000246	1.620	1.042	−0.361
5144	0.800	4.709e−05	0.365	0.940	−0.111	0.800	0.000214	1.659	1.067	−0.218
6000	0.800	4.599e−05	0.404	1.038	−0.094	0.800	0.000193	1.695	1.090	−0.291
	$\eta = 0.8, \alpha = 0.4, \sigma = 4, a_5 = -6000$					$\eta = 0.8, \alpha = 0.4, \sigma = 8, a_5 = -6000$				
99	0.792	0.002284	0.0298	1.302	−1.349	0.769	0.01754	0.229	2.499	−2.190
512	0.796	0.001121	0.0283	1.237	−0.974	0.783	0.00783	0.198	2.160	−2.012
997	0.797	0.000714	0.0235	1.029	−0.525	0.787	0.00630	0.208	2.270	−2.439
2509	0.799	0.000513	0.0245	1.069	−0.387	0.793	0.00293	0.140	1.528	−1.727
3080	0.799	0.000450	0.0233	1.018	−0.453	0.793	0.00303	0.157	1.714	−2.120
4190	0.798	0.000408	0.0239	1.045	−0.454	0.796	0.00205	0.120	1.312	−1.211
5144	0.800	0.000356	0.0226	0.988	−0.473	0.797	0.00182	0.116	1.263	−1.069
6000	0.798	0.000351	0.0237	1.037	−0.283	0.795	0.00174	0.118	1.285	−0.880

bee. The antennal lobe is the primary olfactory structure of the bee brain. It has been demonstrated that stimuli (odors) lead to characteristic activity patterns across spherical functional units, the so-called glomeruli, which collect the converging axonal input from a uniform family of receptor cells. The exact characterization of these patterns is not fully understood and part of an ongoing research program. An important step in the modeling process consists of estimating the response function μ for each individual glomerulus.

Fig. 1(a) shows a typical series, starting 0.2 s after the stimulus and measured for 24 s at time intervals of 0.04 s (i.e. $n = 600$). The fitted curve is obtained by quadratic spline regression with one free knot. The location of the estimated knot is indicated by a dotted vertical line. It may be interpreted as the approximate time where the effect of the stimulus ceases to be noticeable. The residuals appear to be normally distributed in good approximation (Fig. 1(c)) so that we may assume $m = 1$. The log–log-periodogram in Fig. 1(d) shows a clear negative slope. Fitting fractional autoregressive processes [31,34] by maximum likelihood together with model choice based on the BIC (see e.g. [3]) yields the autoregressive order $\hat{p} = 1$. The estimated value of d together with a 95%-confidence interval is equal to 0.219 ± 0.177, and for the autoregressive parameter we have $\hat{\varphi}_1 = 0.481 \pm 0.198$. Fig. 1(d) shows a good agreement between the fitted spectral density and the periodogram. The estimated regression coefficients are $\hat{a}_1 = 947.3 \pm 5.9$, $\hat{a}_2 = -623.5 \pm 158.9$, $\hat{a}_3 = 2196.9 \pm 928.3$ and $\hat{a}_4 = -2223.4 \pm 923.5$. Note in particular that a_4 is clearly different from zero with a p-value of $2.35 \cdot 10^{-6}$, thus confirming the visual impression of a (smooth) change in the trend function. The estimate and 95%-confidence interval for η is equal to 0.147 ± 0.030. In real time this correspond to 3.52 ± 0.71 s. Thus, adding the initial period of 0.2 s not included in our data, the effect of the stimulus is estimated to last for about 3.72 ± 0.71 s.

6. Concluding remarks

In this paper we considered spline regression with residuals exhibiting long-range dependence. The problem was motivated by a question from neurobiology where a structural change in the regression function together with a smooth

Table 2
Simulated mean, variance and skewness of $\hat{a}_5$.

n	Mean	Variance V_{sim}	$V_{sim}/\gamma(N)$	$V_{sim}/[\gamma(N)V_{asym}]$	Skewness	Mean	Variance V_{sim}	$V_{sim}/\gamma(N)$	$V_{sim}/[\gamma(N)V_{asym}]$	Skewness
	$\eta = 0.5, \alpha = 0.8, \sigma = 4, a_5 = -1500$					$\eta = 0.5, \alpha = 0.8, \sigma = 8, a_5 = -1500$				
99	−1558	32 029	10 455 227	1.125	−0.359	−2321	72 195 652	23 567 050 490	634.075	−27.585
512	−1518	8 449	10 335 842	1.112	0.045	−1569	41 203	50 406 411	1.356	−1.167
997	−1509	4528	9 447 317	1.017	−0.099	−1538	22 145	46 206 401	1.243	−0.353
2509	−1504	2 014	8 795 468	0.947	0.02	−1516	8 713	38 058 186	1.024	−0.012
3080	−1503	1 798	9 255 180	0.996	0.003	−1514	7 399	38 080 936	1.025	−0.049
4190	−1503	1 438	9 468 399	1.019	0.020	−1512	5 570	36 675 250	0.987	0.038
5144	−1504	1 146	8 892 199	0.957	−0.075	−1509	4 734	36 733 482	0.988	−0.002
6000	−1501	1 081	9 482 803	1.021	−0.041	−1508	4 039	35 449 080	0.954	0.054
	$\eta = 0.5, \alpha = 0.4, \sigma = 4, a_5 = -1500$					$\eta = 0.5, \alpha = 0.4, \sigma = 8, a_5 = -1500$				
99	−1607	67 620	881 700	1.337	−0.801	−19 337	325 265 371 462	4 241 162 495 276	608 367.254	−35.865
512	−1554	29 972	756 563	1.148	−0.168	−1 830	952 124	24 033 684	9.114	−13.774
997	−1537	21 692	715 107	1.085	−0.109	−1 753	970 002	31 976 809	12.126	−19.897
2509	−1526	13 613	649 316	0.985	−0.007	−1 641	84 812	4 045 243	1.534	−1.602
3080	−1524	13 418	694 735	1.054	0.028	−1 618	102 163	5 289 513	2.006	−4.824
4190	−1521	11 545	676 052	1.026	−0.021	−1 597	96 487	5 650 302	2.143	−7.501
5144	−1522	10 287	653 926	0.992	0.079	−1 582	58 105	3 693 685	1.401	−0.826
6000	−1515	10 250	692 969	1.051	0.039	−1 581	48 709	3 293 045	1.249	−0.392
	$\eta = 0.8, \alpha = 0.8, \sigma = 4, a_5 = -6000$					$\eta = 0.8, \alpha = 0.8, \sigma = 8, a_5 = -6000$				
99	−6784	11 792 276	3 849 389 314	1.555	−2.709	−35 854	423 050 639 425	138 097 731 229 069	13 950.215	−34.848
512	−6222	2 128 106	2 603 445 507	1.052	−0.819	−6 962	18 110 572	22 155 800 179	2.238	−4.122
997	−6156	1 311 885	2 737 296 542	1.106	−0.582	−6 441	6 718 153	14 017 676 417	1.416	−1.558
2509	−6064	572 699	2 501 539 158	1.011	−0.382	−6 307	2 560 854	11 185 760 648	1.130	−0.645
3080	−6059	526 958	2 712 206 529	1.096	−0.396	−6 170	2 185 482	11 248 477 157	1.136	−0.721
4190	−6028	391 480	2 577 608 740	1.042	−0.258	−6 188	1 661 868	10 942 174 482	1.105	−0.668
5144	−6046	307 060	2 382 402 539	0.963	−0.289	−6 186	1 429 679	11 092 527 358	1.121	−0.639
6000	−6047	294 209	2 581 888 324	1.043	−0.328	−6 111	1 243 206	10 910 002 939	1.102	−0.543
	$\eta = 0.8, \alpha = 0.4, \sigma = 4, a_5 = -6000$					$\eta = 0.8, \alpha = 0.4, \sigma = 8, a_5 = -6000$				
99	−6782	14 121 482	184 131 184	1.385	−1.961	−45 863	642 758 339 288	8 380 979 966 787	15 760.494	−27.728
512	−6420	7 286 332	183 922 987	1.383	−1.789	−9 225	124 150 310	3 133 825 821	5.893	−5.001
997	−6347	4 600 611	151 662 423	1.141	−1.295	−8 602	113 295 891	3 734 879 852	7.023	−6.724
2509	−6249	3 184 762	151 903 155	1.143	−0.968	−7 485	33 286 751	1 587 673 463	2.986	−4.163
3080	−6266	2 861 525	148 156 304	1.114	−0.869	−7 443	41 830 206	2 165 771 515	4.073	−12.001
4190	−6172	2 407 519	140 984 920	1.06	−0.607	−7 039	15 868 524	929 264 703	1.747	−2.086
5144	−6233	2 146 698	136 463 432	1.026	−0.533	−7 140	15 162 270	963 850 325	1.813	−1.979
6000	−6127	2 166 628	146 478 766	1.102	−0.890	−6 893	13 702 442	926 378 109	1.742	−2.042

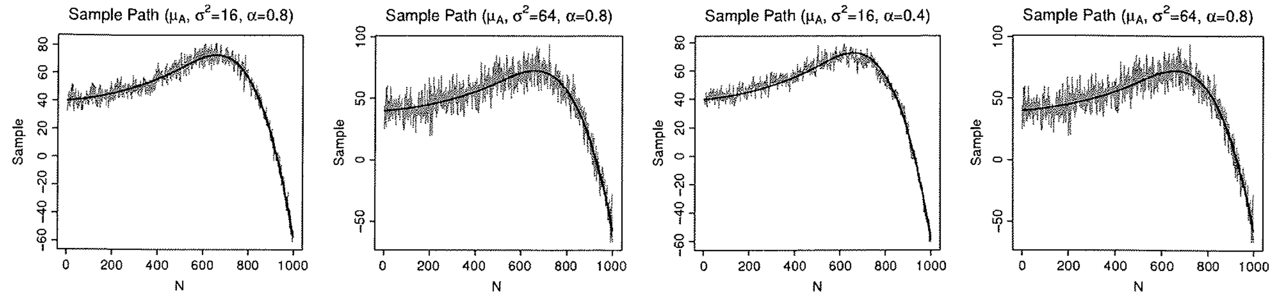

Fig. 3. Observed sample paths with trend function μ_A ($\eta = 0.5, a_5 = -1500$), $\alpha \in \{0.4, 0.8\}$ and $\sigma^2 \in \{16, 64\}$.

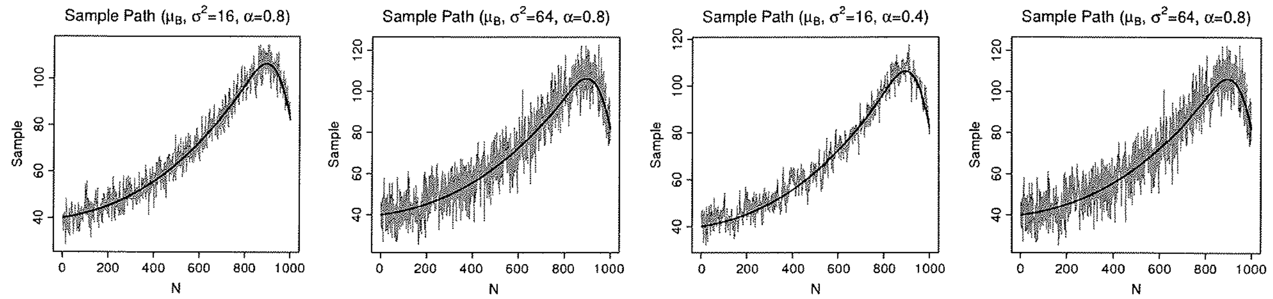

Fig. 4. Observed sample signal with trend function μ_B ($\eta = 0.8, a_5 = -6000$)$\alpha \in \{0.4, 0.8\}$ and $\sigma^2 \in \{16, 64\}$.

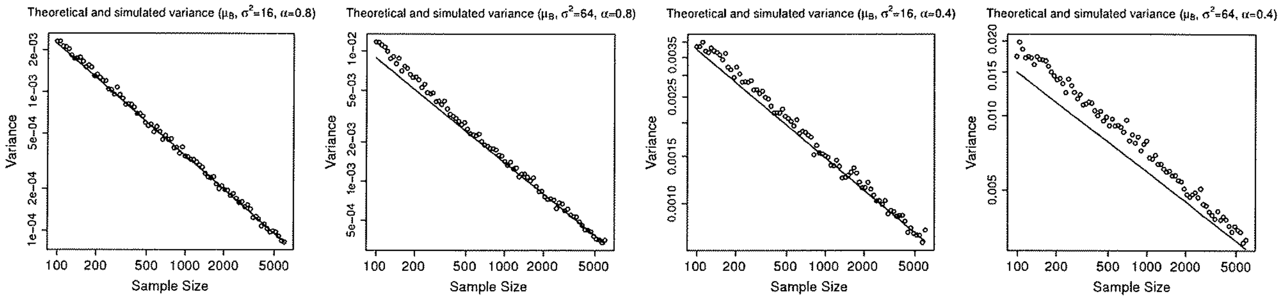

Fig. 5. Log–log-plots of $\text{Var}_{\text{sim}}(\hat{\eta})$ for the trend function μ_B ($\eta = 0.5, a_5 = -1500$), $\alpha \in \{0.4, 0.8\}$ and $\sigma^2 \in \{16, 64\}$.

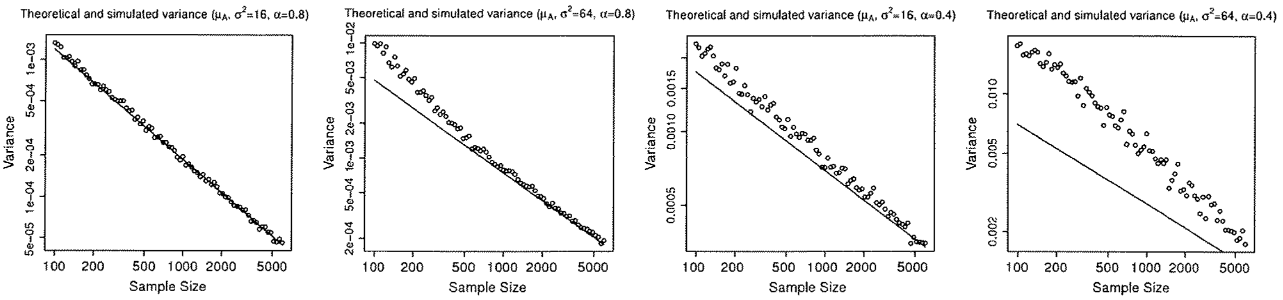

Fig. 6. Log–log-plots of $\text{Var}_{\text{sim}}(\hat{\eta})$ for the trend function μ_B ($\eta = 0.8, a_5 = -6000$), $\alpha \in \{0.4, 0.8\}$ and $\sigma^2 \in \{16, 64\}$.

transition between the two phases is known to occur. As expected, (strong) long memory changes the rate of convergence of parameter estimates. A useful extension that will need to be looked at is the case of antipersistence, i.e. $\sum \gamma_\varepsilon(k) = 0$. For instance, in the neurobiological experiment described above, not only long memory but also antipersistence can be observed for many of the glomeruli included in the study.

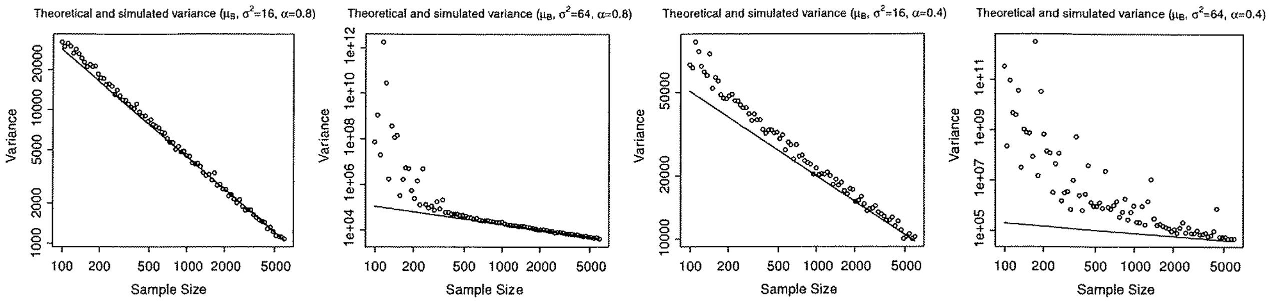

Fig. 7. Log–log-plots of $\mathrm{Var}_{\mathrm{sim}}(\hat{a}_5)$ for the trend function μ_A ($\eta = 0.5, a_5 = -1500$). $\alpha \in \{0.4, 0.8\}$ and $\sigma^2 \in \{16, 64\}$.

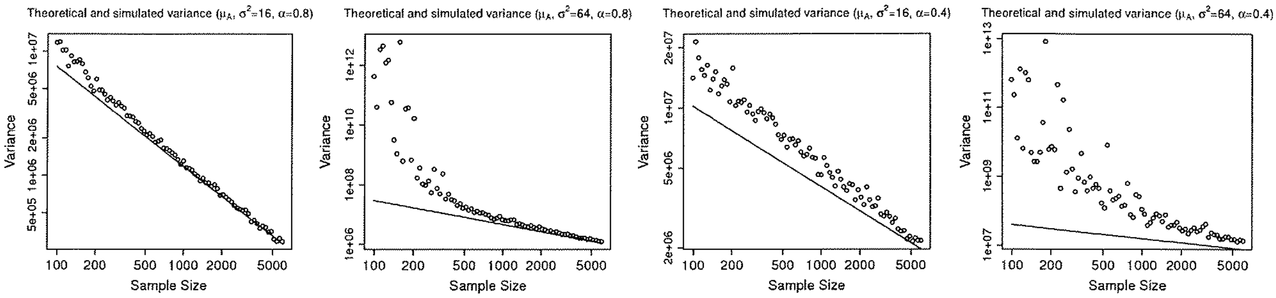

Fig. 8. Log–log-plots of $\mathrm{Var}_{\mathrm{sim}}(\hat{a}_5)$ for the trend function μ_B ($\eta = 0.8, a_5 = -6000$), $\alpha \in \{0.4, 0.8\}$ and $\sigma^2 \in \{16, 64\}$.

Acknowledgments

We would like to thank Giovanni Galizia and Martin Strauch (Department of Biology, University of Konstanz) for providing us with the data and for valuable discussions regarding the biological context. We would also like to thank the referees for many valuable suggestions.

Appendix

For simplicity of presentation, all proofs are first formulated for cubic splines. The proofs carry over directly to higher order splines without any change (except for replacing specific notations with $p = 3$ to general p). The modifications needed to adapt the proofs to quadratic splines ($p = 2$) are given in Appendix A.3.

A.1. Proofs of Section 3.1

The following lemma is taken from [28, page 9]. It is a direct consequence of a more general statement in [11].

Lemma 2. *Let $A \subset \mathbb{R}^k$ and $B \subset \mathbb{R}^n$ be two compact sets. Let $g(a, b)$ denote a real valued, continuous function defined on $A \times B$. If ν_l denotes a sequence of probability measures on A converging weakly to a probability measure ν on A, then*

$$\int_A g(a, b)\,\mathrm{d}\nu_l(a) \longrightarrow \int_A g(a, b)\,\mathrm{d}\nu(a)$$

uniformly in b.

Definition 1. Let $\eta \in (0, 1)$. We define the matrices

$$\mathbf{V_{1,n}}(\eta) = \left[(t/n)^{j-1}\mathbf{1}_{(0,\eta]}(t/n)\right]_{t=1,\ldots,n;j=1,\ldots,4} \in \mathbb{R}^{n\times 4}$$

and

$$\mathbf{V_{2,n}}(\eta) = \left[(t/n)^{j-1}\mathbf{1}_{(\eta,1]}(t/n)\right]_{t=1,\ldots,n;j=1,\ldots,4} \in \mathbb{R}^{n\times 4}.$$

Let $0 < a < b \le 1$ such that $q = \lfloor bn \rfloor - \lfloor an \rfloor \ge 1$. We define the matrix

$$\mathbf{U_n}(a, b) = \left[\delta_{tj}\right]_{t=1,\ldots,n;j=\lfloor an\rfloor+1,\ldots,\lfloor bn\rfloor} \in \mathbb{R}^{n\times q}.$$

Remark 3. Note that $\lfloor an \rfloor + 1 \le j \le \lfloor bn \rfloor$ if and only if $a < j/n \le b$.

Remark 4. The column spaces of $\mathbf{V_{1,n}}(\eta)$ and $\mathbf{V_{1,n}}(\eta)$ are orthogonal. If $a < b$ then the column spaces of $\mathbf{V_{1,n}}(a)$, $\mathbf{V_{1,n}}(b)$ and $\mathbf{U_n}(a, b)$ are orthogonal.

Proof of Lemma 1. We will prove the lemma by contradiction. Assume that for some $\Delta > 0$

$$\liminf_n \left(\inf_{u \in K(\Delta)} n^{-1} \Phi_n (\theta, \theta + u) \right) = 0.$$

Then there exists a sequence $(n_k)_{k\in\mathbb{N}}$ such that

$$\left(\inf_{u \in K(\Delta)} n_k^{-1} \Phi_{n_k} (\theta, \theta + u) \right) \leq \frac{1}{2 \cdot k}.$$

In addition, for each $k \in \mathbb{N}$ there exists $u_k \in K(\Delta)$, such that

$$\left| \inf_{u \in K(\Delta)} n_k^{-1} \Phi_{n_k} (\theta, \theta + u) - n_k^{-1} \Phi_{n_k} (\theta, \theta + u_k) \right| \leq \frac{1}{2 \cdot k}.$$

The triangle inequality implies $\left| n_k^{-1} \Phi_{n_k} (\theta, \theta + u_k) \right| \leq \frac{1}{k}$, i.e. we can find a sequence $(\theta_k)_{k\in\mathbb{N}}, \theta_k = (a_k, \eta_k)$ such that $\|\theta - \theta_k\|^2 > \Delta$ for all k and

$$\lim_k n_k^{-1} \Phi_{n_k} (\theta, \theta_k) = 0.$$

Since $\eta_k \in (0, 1)$, we may assume (by choosing a subsequence) that $\lim_k \eta_k = \eta_\infty \in [0, 1]$. Because $\eta \in (0, 1)$, we can find $\Delta^* \leq \frac{\Delta}{2}$ such that $[\eta - \Delta^*, \eta + \Delta^*] \subset (0, 1)$.

Case 1: $|\eta_\infty - \eta|^2 \leq \Delta^*$. In this case we may assume that the sequence $(a_k)_{k\in\mathbb{N}}$ is bounded: If $\|a_k\|^2 \to \infty$ then

$$\begin{aligned} n_k^{-1} \left\| \mu_{n_k}(\theta_k) \right\|^2 &= n_k^{-1} a_k' \mathbf{W}_{\mathbf{n_k}}(\eta_k)' \mathbf{W}_{\mathbf{n_k}}(\eta_k) a_k \\ &\geq \|a_k\|^2 \lambda_{\min} \left(n_k^{-1} \mathbf{W}_{\mathbf{n_k}}(\eta_k)' \mathbf{W}_{\mathbf{n_k}}(\eta_k) \right) \to \infty. \end{aligned}$$

Note that $\lambda_{\min} \left(n_k^{-1} \mathbf{W}_{\mathbf{n_k}}(\eta_k)' \mathbf{W}_{\mathbf{n_k}}(\eta_k) \right)$ is bounded away from zero for large k, since η_k is bounded away from both 0 and 1. We conclude

$$\begin{aligned} n_k^{-1} \Phi (\theta, \theta_k) &= n_k^{-1} \left\| \mu_{n_k}(\theta) \right\|^2 - 2 \langle \mu_{n_k}(\theta), \mu_{n_k}(\theta) \rangle + n_k^{-1} \left\| \mu_{n_k}(\theta_k) \right\|^2 \\ &\geq n_k^{-1} \left\| \mu_{n_k}(\theta) \right\|^2 + n_k^{-1} \left\| \mu_{n_k}(\theta_k) \right\| \left(\left\| \mu_{n_k}(\theta_k) \right\| - \left\| \mu_{n_k}(\theta) \right\| \right) \to \infty, \end{aligned}$$

a contradiction to $n_k^{-1} \Phi (\theta, \theta_k) \to 0$. If a_k is bounded, however, we may (by choosing a subsequence) assume that $a_k \to a_\infty$. Since $|\eta_\infty - \eta|^2 \leq \Delta^*$, we have $\|a_\infty - a\|^2 \geq \Delta/2$, i.e. we found a sequence $(\theta_k)_{k\in\mathbb{N}}$ such that $\lim_k \theta_k = \theta_\infty \neq \theta$ and $n_k^{-1} \Phi_{n_k}(\theta, \theta_k) \to 0$. By Lemma 2

$$\int_0^1 |\mu (t, \theta) - \mu (t, \theta_\infty)|^2 \, dt = 0,$$

a contradiction to $\theta_\infty \neq \theta$.

Case 2: $|\eta_\infty - \eta|^2 > \Delta^*$. The properties of the orthogonal projection imply

$$n_k^{-1} \left\| \mu_{n_k}(\theta) \right\|^2 - n_k^{-1} \left\| P_{\mathbf{W}_{\mathbf{n_k}}(\eta_k)} \mu_{n_k}(\theta) \right\|^2 = n_k^{-1} \left\| \mu_k(\theta) - P_{\mathbf{W}_{\mathbf{n_k}}(\eta_k)} \mu_{n_k}(\theta) \right\|^2 \leq n_k^{-1} \Phi_{n_k} (\theta, \theta_k) \to 0. \tag{10}$$

If $\eta_\infty \in (0, 1)$, $\mathbf{W}_{\mathbf{n_k}}(\eta_k)$ has full rank for k large enough. In this case

$$P_{\mathbf{W}_{\mathbf{n_k}}(\eta_k)} \mu_{n_k} = \mathbf{W}_{\mathbf{n_k}}(\eta_k) a_k$$

where $a_k = \left(\mathbf{W}_{\mathbf{n_k}}(\eta_k)' \mathbf{W}_{\mathbf{n_k}}(\eta_k) \right)^{-1} \mathbf{W}_{\mathbf{n_k}}(\eta_k)' \mu_{n_k}(\theta)$. Lemma 2 implies $\lim_k a_k = a_\infty$ where a_∞ denotes the regression coefficients of $\mu(s, \theta)$ on the space $\mathbf{W}_\infty = \text{span} \left(s^0, \dots, s^p, (s - \eta_\infty)_+^p \right)$ within $L^2([0, 1])$. Applying Lemma 2 once more shows that

$$n_k^{-1} \left\| \mathbf{W}_{\mathbf{n_k}}(\eta_{n_k}) a_k \right\|^2 \longrightarrow \left\| P_{\mathbf{W}_\infty} \mu (\cdot, \theta) \right\|^2_{L^2([0,1])},$$

a contradiction to (10). Assume now that $\eta_{n_k} \to 1$. Then

$$n_k^{-1} \left\| P_{\mathbf{W}_{\mathbf{n}}(\eta_{n_k})} \mu_n(\theta) \right\|^2 \leq n_k^{-1} \left\| P_{\mathbf{V}_{\mathbf{1,n}}(\eta_{n_k})} \mu_n(\theta) \right\|^2 + n_k^{-1} \left\| P_{\mathbf{V}_{\mathbf{2,n}}(\eta_{n_k})} \mu_n(\theta) \right\|^2.$$

Since

$$n_k^{-1} \left\| P_{\mathbf{V}_{\mathbf{2,n}}(\eta_{n_k})} \mu_n(\theta) \right\|^2 \leq n_k^{-1} \sum_{t=\lfloor \eta_{n_k} n_k \rfloor}^{n_k} \mu^2 (t/n_k, \theta) \to 0$$

and

$$\lim_k n_k^{-1} \left\| P_{\mathbf{V_{1,n}}(\eta_{n_k})} \mu_n(\theta) \right\|^2 = \left\| P_{\mathbf{V}_\infty} \mu(\cdot, \theta) \right\|^2_{L^2([0,1])}$$

where $\mathbf{V}_\infty = \text{span}\left(s^0, \ldots, s^p\right)$, we again obtain a contradiction to (10). Exchanging the roles of $\mathbf{V_{1,n}}(\eta_{n_k})$ and $\mathbf{V_{2,n}}(\eta_{n_k})$ one can treat case $\eta_{n_k} \to 0$. □

Theorem 1 follows from the following Lemmas 3 and 4.

Lemma 3. *Let η be a fixed knot. Then*

$$\mathbb{P}\left(\frac{1}{n}\left\langle \mathbf{e}_n, P_{\mathbf{W_n}(\eta)}\mathbf{e}_n\right\rangle > \Delta\right) = O(\gamma^m(n)). \tag{11}$$

Proof of Lemma 3. Since $P_{\mathbf{W_n}}$ is a projection matrix on the five-dimensional vector space span($\mathbf{W_n}$), it can be written as

$$P_{\mathbf{W_n}} = \mathbf{A_n'D_nA_n}$$

with $\mathbf{A_n} = (\mathbf{a}_{kj;n}) \in \mathbb{R}^{n\times n}$ an orthogonal matrix and $\mathbf{D_n} \in \mathbb{R}^{n\times n}$ a diagonal matrix with

$$d_{ii} = \begin{cases} 1 & i = 1, \ldots, 5 \\ 0 & i = 6, \ldots, n. \end{cases}$$

Denoting by $\mathbf{a}_{k\cdot;n}$ the kth row of $\mathbf{A}_n$, interpreted as a column vector, we have $(\mathbf{A}_n\mathbf{e}_n)_k = \sum_{j=0}^n \mathbf{a}_{kj;n}\xi_j = \left\langle \mathbf{a}_{k\cdot;n}, \mathbf{e}_n\right\rangle$ so that

$$\frac{1}{n}\mathbf{e}_n' P_{\mathbf{W_n}(\eta)}\mathbf{e}_n = \frac{1}{n}(\mathbf{A}_n\mathbf{e}_n)'\mathbf{D_n}\,(\mathbf{A}_n\mathbf{e}_n) = \sum_{k=1}^{5} \frac{1}{\sqrt{n}}\left\langle \mathbf{a}_{k\cdot;n}, \mathbf{e}_n\right\rangle \cdot \frac{1}{\sqrt{n}}\left\langle \mathbf{a}_{k\cdot;n}, \mathbf{e}_n\right\rangle. \tag{12}$$

Since $\mathbb{E}\left[\left\langle \mathbf{a}_{k\cdot;n}, \mathbf{e}_n\right\rangle\right] = 0$, it is sufficient to show that $\text{Var}\left(n^{-\frac{1}{2}}\left\langle \mathbf{a}_{k\cdot;n}, \mathbf{e}_n\right\rangle\right) = O(\gamma^m(n))$: Let $\Sigma_\mathbf{e} = \mathbf{U'SU}$ denote the covariance matrix of $\mathbf{e}_n$ and its corresponding spectral decomposition. Noting that $\|\mathbf{U}\mathbf{a}_{k\cdot;n}\| = 1$, we have

$$\text{Var}\left(n^{-\frac{1}{2}}\left\langle \mathbf{a}_{k\cdot;n}, \mathbf{e}_n\right\rangle\right) = \frac{1}{n}(\mathbf{U}\mathbf{a}_{k\cdot;n})'\mathbf{S}(\mathbf{U}\mathbf{a}_{k\cdot;n}) \le \frac{1}{n}\lambda_{\max}(\mathbf{S}) = \frac{1}{n}\lambda_{\max}(\Sigma_\mathbf{e}).$$

Since $\gamma_\xi(k) \sim \frac{C_m^2}{m!}\gamma^m(k)$, we can conclude

$$\begin{aligned} n^{-1}\lambda_{\max}(\Sigma_\mathbf{e}) &= n^{-1}\max_{\|\nu\|=1} \nu'\Sigma_e\nu \le n^{-1}\sum_{k=-n+1}^{n-1}\left|\gamma_\xi(k)\right|\sum_i |\nu_i|\,|\nu_{i+k}| \\ &\le n^{-1}\sum_{k=-n+1}^{n-1}\left|\gamma_\xi(k)\right| = O(\gamma^m(n)). \quad \square \end{aligned}$$

A stronger result states that Eq. (11) holds uniformly in $\eta \in (0, 1)$:

Lemma 4. *Under the same assumptions as in Lemma 3 we have,*

$$\mathbb{P}\left(\frac{1}{n}\sup_{\tilde\eta\in(0,1)}\left\langle \mathbf{e}_n, P_{\mathbf{W_n}(\tilde\eta)}\mathbf{e}_n\right\rangle > \Delta\right) = O(\gamma^\kappa(n)) \tag{13}$$

where $\kappa = \min(j, m)$.

Proof of Lemma 4. Without loss of generality we may assume $\text{Var}(\xi_i) = 1$. For $\Delta > 0$, we can choose a finite sequence $(\eta_j)_{j=1}^M$ such that $\cup_{j=1}^M B(\eta_j, \Delta/10) = [0, 1]$ where $B(\eta_j, \Delta/10) = B_j$ $(j = 1, 2, \ldots, M)$ denote closed $\Delta/10$-balls around η_j. Then

$$\mathbb{P}\left(\frac{1}{n}\sup_{\tilde\eta\in(0,1)}\left\langle \mathbf{e}_n, P_{\mathbf{W_n}(\tilde\eta)}\mathbf{e}_n\right\rangle > \Delta\right) \le \sum_{j=1}^M \underbrace{\mathbb{P}\left(\frac{1}{n}\sup_{\tilde\eta\in B_j}\left\langle \mathbf{e}_n, P_{\mathbf{W_n}(\tilde\eta)}\mathbf{e}_n\right\rangle > \Delta\right)}_{=:C_j}.$$

It is thus sufficient to show convergence of C_j to zero. Without loss of generality, we may assume $x_j^* = \inf\left\{x : x \in B_j\right\} \notin \mathbb{Q}$ and $y_j^* = \sup\left\{y : y \in B_j\right\} \notin \mathbb{Q}$.

Let $P_{\mathbf{V_{1,n}},(x_j^*),\mathbf{V_{2,n}}(y_j^*)}$ denote the orthogonal projection on the space span $\left(\mathbf{V_{1,n}}\left(x_j^*\right)\right) \oplus$ span $\left(\mathbf{V_{2,n}}(y_j^*)\right)$. Since

$$\text{span}(\mathbf{W_n}(\tilde{\eta})) \subset \text{span}\left(\mathbf{V_{1,n}}\left(x_j^*\right)\right) \oplus \text{span}\left(\mathbf{V_{2,n}}(y_j^*)\right) \oplus \text{span}\left(\mathbf{U_n}\left(x_j^*, y_j^*\right)\right)$$

for all $\tilde{\eta} \in B_j$, we have

$$\mathbb{P}\left(\sup_{\tilde{\eta}\in B_j} \frac{1}{n}\left\langle \mathbf{e}_n, P_{\mathbf{W_n}(\tilde{\eta})}\mathbf{e}_n\right\rangle > \Delta\right) \le \mathbb{P}\left(\frac{1}{n}\left\langle \mathbf{e}_n, P_{\mathbf{V_{1,n}}\left(x_j^*\right),\mathbf{V_{2,n}}(y_j^*)}\mathbf{e}_n\right\rangle > \frac{\Delta}{2}\right) + \mathbb{P}\left(\frac{1}{n}\left\langle \mathbf{e}_n, P_{\mathbf{U_n}\left(x_j^*,y_j^*\right)}\mathbf{e}_n\right\rangle > \frac{\Delta}{2}\right).$$

From

$$\begin{aligned}\Xi &:= \mathbb{E}\left(\frac{1}{n}\left\langle \mathbf{e}_n, P_{\mathbf{U_n}\left(x_j^*,y_j^*\right)}\mathbf{e}_n\right\rangle\right) = \mathbb{E}\left(\frac{1}{n}\sum_{k=\lfloor x_j^* n\rfloor+1}^{\lfloor y_j^* n\rfloor} \xi_k^2\right)\\ &\le \frac{1+n\Delta/5}{n}\text{Var}(\xi_1) \le \frac{\Delta}{4}\end{aligned}$$

for sufficiently large n, we can conclude again by Chebyshev's inequality

$$\mathbb{P}\left(\frac{1}{n}\left\langle \mathbf{e}_n, P_{\mathbf{U_n}\left(x_j^*,y_j^*\right)}\mathbf{e}_n\right\rangle > \frac{\Delta}{2}\right) \le \text{Var}\left(\frac{1}{n}\sum_{k=\lfloor x_j^* n\rfloor+1}^{\lfloor y_j^* n\rfloor} \xi_k^2\right)\left(\frac{\Delta}{2} - \Xi\right)^{-2} = O(\gamma^j(n)).$$

Since the dimension of span $\left(\mathbf{V_{1,n}}\left(x_j^*\right)\right) \oplus$ span $\left(\mathbf{V_{2,n}}(y_j^*)\right)$ is always equal to 8 (for sufficiently large n), a modification of Lemma 3 implies

$$\mathbb{P}\left(\frac{1}{n}\left\langle \mathbf{e}_n, P_{\mathbf{V_{1,n}},\left(x_j^*\right),\mathbf{V_{2,n}}(y_j^*)}\mathbf{e}_n\right\rangle > \frac{\Delta}{2}\right) = O(\gamma^\kappa(n)).$$

This concludes the proof. □

Proof of Theorem 1. By the definition of the LSE, we have

$$\begin{aligned}\sum_{t=1}^n \xi_t^2 &= \sum_{t=1}^n [\xi_t + \mu(t/n,\theta) - \mu(t/n,\theta)]^2 \ge \sum_{t=1}^n \left[\xi_t + \mu(t/n,\theta) - \mu\left(t/n,\hat{\theta}\right)\right]^2\\ &= \sum_{t=1}^n \xi_t^2 + \Phi_n(\theta,\hat{\theta}) + 2\sum_{t=1}^n \xi_t\left[\mu(t/n,\theta) - \mu\left(t/n,\hat{\theta}\right)\right],\end{aligned}$$

and hence

$$\frac{1}{2} \le \frac{\sum_{t=1}^n \xi_t\left[\mu(t/n,\hat{\theta}) - \mu(t/n,\theta)\right]}{\Phi_n(\theta,\hat{\theta})}.$$

Let $|\theta - \hat{\theta}| > \Delta$. According to Lemma 1 there exists $N_\Delta \in \mathbb{N}$ such that $\Phi_n(\theta,\hat{\theta}) \ge \delta n$ for all $n \ge N_\Delta$ and a suitable $\delta > 0$. Therefore

$$\begin{aligned}\mathbb{P}(|\hat{\theta}_n - \theta| \ge \Delta) &\le \mathbb{P}\left(\frac{\delta}{2} \le \frac{1}{n}\sum_{t=1}^n \xi_t\left[\mu(t/n,\hat{\theta}) - \mu(t/n,\theta)\right]\right)\\ &= \mathbb{P}\left(\frac{\delta}{2} \le \frac{\left\langle \mathbf{e}_n, P_{\mathbf{W_n}(\hat{\eta})}(\mu_n(\theta)+\mathbf{e}_n) - \mu_n(\theta)\right\rangle}{n}\right)\\ &\le \mathbb{P}\left(\frac{\delta}{6} \le \frac{\left|\left\langle \mathbf{e}_n, P_{\mathbf{W_n}(\hat{\eta})}\mathbf{e}_n\right\rangle\right|}{n}\right) + \mathbb{P}\left(\frac{\delta}{6} \le \frac{\left|\left\langle \mathbf{e}_n, P_{\mathbf{W_n}(\hat{\eta})}\mu_n(\theta)\right\rangle\right|}{n}\right) + \mathbb{P}\left(\frac{\delta}{6} \le \frac{|\langle \mathbf{e}_n, \mu_n(\theta)\rangle|}{n}\right).\end{aligned} \tag{14}$$

The first term in (14) vanishes according to Lemma 4. Moreover, since $\mathbb{E}\left(n^{-1}\langle \mathbf{e}_n, \mu_n(\theta)\rangle\right) = 0$ and

$$\begin{aligned}\text{Var}\left(\frac{\langle \mathbf{e}_n, \mu_n(\theta)\rangle}{n}\right) &= \frac{1}{n^2}\sum_{s,t=1}^n \text{Cov}(\xi_s,\xi_t)\mu(t/n,\theta)\mu(s/n,\theta)\\ &\le \mathbf{const}\frac{1}{n^2}\sum_{s,t=1}^n |\text{Cov}(\xi_s,\xi_t)|,\end{aligned}$$

the third term in (14) vanishes asymptotically. In order to see that the second term converges to zero, let $v_{k,n}(\hat{\eta}), k = 1, \ldots, 5$ denote an orthonormal basis of span$(\mathbf{W_n}(\hat{\eta}))$ (with $v_{1,n}, \ldots, v_{4,n}$ not depending on $\hat{\eta}$). Then

$$\frac{1}{n}\left\langle \mathbf{e}_n, P_{\mathbf{W_n}(\hat{\eta})}\mu_n(\theta)\right\rangle = \frac{1}{n}\sum_{j=1}^{4}\left\langle \mathbf{e}_n, P_{v_{j,n}}\mu_n(\theta)\right\rangle + \frac{1}{n}\left\langle \mathbf{e}_n, P_{\mathbf{v}_{k,5}(\hat{\eta})}\mu_n(\theta)\right\rangle.$$

The first term does not depend on $\hat{\eta}$. Hence, by the same calculations as before, it converges to zero in probability. For the second term

$$\frac{1}{n}\left\langle \mathbf{e}_n, P_{\mathbf{v}_{k,5}(\hat{\eta})}\mu_n(\theta)\right\rangle = \frac{1}{\sqrt{n}}\left\langle \mathbf{e}_n, \mathbf{v}_{k,5}(\hat{\eta})\right\rangle \cdot \frac{1}{\sqrt{n}}\left\langle \mu_n(\theta), \mathbf{v}_{k,5}(\hat{\eta})\right\rangle$$

with $n^{-\frac{1}{2}}\left|\left\langle \mu_n(\theta), \mathbf{v}_{k,5}(\hat{\eta})\right\rangle\right| \leq n^{-\frac{1}{2}}\|\mu_n(\theta)\|$ being uniformly bounded. Now

$$\left|n^{-\frac{1}{2}}\left\langle \mathbf{e}_n, \mathbf{v}_{k,5}(\hat{\eta})\right\rangle\right| = n^{-\frac{1}{2}}\left\|P_{\mathbf{v}_{k,5}(\hat{\eta})}\mathbf{e}_n,\right\| \leq n^{-\frac{1}{2}}\left\|P_{\mathbf{W_n}(\hat{\eta})}\mathbf{e}_n,\right\|$$

and so by Lemma 4

$$\mathbb{P}\left(n^{-1}\left|\left\langle \mathbf{e}_n, P_{\mathbf{v}_{k,5}(\hat{\eta})}\mu_n(\theta)\right\rangle\right| > \Delta\right) = O\left(\gamma^{\kappa}(n)\right).$$

Finally, the exact order of $\mathbb{P}(|\hat{\theta}_n - \theta| \geq \Delta)$ follows from the upper bounds given here, applying the asymptotic formulas

$$\text{Var}\left(\frac{1}{n}\sum_{t=1}^{n} G(\varepsilon_t)\right) = O(\gamma^m(n)), \qquad \text{Var}\left(\frac{1}{n}\sum_{t=1}^{n} G^2(\varepsilon_t)\right) = O(\gamma^j(n)). \quad \square$$

A.2. Proofs of Section 3.2

The proof Theorem 2 requires some auxiliary results which we need to establish prior to the main proof.

Lemma 5. *Let $f : [0, 1] \times [0, 1] \longrightarrow \mathbb{R}_+$ be a continuous function and let $\gamma(t-s) = L(t-s)|t-s|^{-\alpha}$ denote an autocorrelation function. Then*

$$\lim_{n\to\infty} \gamma^{-m}(n)\frac{1}{n^2}\sum_{s,t=1}^{n}\gamma^m(t-s)f\left(\frac{t}{n}, \frac{s}{n}\right) = \int_0^1\int_0^1 f(t,s)|t-s|^{-\alpha m}\,dt\,ds. \tag{15}$$

Proof of Lemma 5. It is sufficient to show that

$$\left|\frac{n^{\alpha m}}{L^m(n)n^2}\sum_{s,t=1}^{n}\gamma^m(t-s)f\left(\frac{t}{n}, \frac{s}{n}\right) - \frac{1}{n^2}\sum_{s,t=1,s\neq t}^{n} f\left(\frac{t}{n}, \frac{s}{n}\right)\left|\frac{t}{n} - \frac{s}{n}\right|^{-\alpha m}\right| \to 0. \tag{16}$$

Let $0 < \delta < 1 - \alpha m$. Define the sequences of functions on $[0, 1] \times [0, 1]$:

$$g_n(u, v) := \sum_{t,s=1,s\neq t}^{n}\mathbf{1}_{\left(\frac{t-1}{n}, \frac{t}{n}\right]}(u)\mathbf{1}_{\left(\frac{s-1}{n}, \frac{s}{n}\right]}(v)\left|\frac{L(t-s)}{L(n)} - 1\right|\left|\frac{t-s}{n}\right|^{\delta}$$

$$f_n(u, v) := \sum_{t,s=1,s\neq t}^{n}\mathbf{1}_{\left(\frac{t-1}{n}, \frac{t}{n}\right]}(u)\mathbf{1}_{\left(\frac{s-1}{n}, \frac{s}{n}\right]}(v)f\left(\frac{t}{n}, \frac{s}{n}\right)\left|\frac{t}{n} - \frac{s}{n}\right|^{-\alpha m-\delta}.$$

Since $\gamma(n) = L(n)n^{-\alpha}$ the left side of (16) is equal to

$$\begin{aligned}
&\left|\frac{1}{n^2}\sum_{s,t=1,s\neq t}^{n}\left[\frac{L^m(t-s)}{L^m(n)} - 1\right]f\left(\frac{t}{n}, \frac{s}{n}\right)\left|\frac{t}{n} - \frac{s}{n}\right|^{-\alpha m}\right| + o(1) \\
&\leq \frac{1}{n^2}\sum_{s,t=1,s\neq t}^{n}\left|f\left(\frac{t}{n}, \frac{s}{n}\right)\right|\left|\frac{L(t-s)^m}{L(n)^m} - 1\right|\left|\frac{t}{n} - \frac{s}{n}\right|^{-\alpha m} + o(1) \\
&= \int_0^1\int_0^1 |f_n(u, v)|g_n(u, v)\,dv\,du + o(1).
\end{aligned}$$

Since $\int_0^1\int_0^1 |f_n(u,v)|\,dv\,du$ converges to $\int_0^1\int_0^1 |f(u,v)|\,|u-v|^{-\alpha m-\delta}dv\,du$, it is sufficient to show that $g_n(u,v) \to 0$ uniformly in $(u,v) \in [0,1]\times[0,1]$. However, with $x=(t-s)/n$, we have

$$\left|\frac{L(t-s)}{L(n)} - 1\right|\left|\frac{t-s}{n}\right|^{\delta} = \left|\frac{L(nx)(nx)^{\delta}}{L(n)n^{\delta}} - x^{\delta}\right| = h_n(x). \tag{17}$$

The function $x \mapsto L(x)x^{\delta}$ is of regular variation. Moreover $\gamma(t-s) = L(t-s)|t-s|^{-\alpha} \le 1$ implies that $L(x)$ is bounded in each interval $(0,a] \subset \mathbb{R}$ for all $a>0$. It then follows from Theorem 1.5.2 on page 22 in [12] that $h_n(x)$ converges uniformly to zero. □

Lemma 6. *Let* $\|\hat{u}\| < \delta$ *where* $\hat{u} = \hat{\theta} - \theta$. *For all* $k = 1,\ldots,6$ *there exists a vector* $\mathbf{r}_{k\cdot;n} \in \mathbb{R}^6$ *such that*

$$\sum_{t=1}^{n} \xi_t\left[\mu_{(k)}(t/n,\theta+\hat{u}) - \mu_{(k)}(t/n,\theta)\right] = \left\langle \mathbf{r}_{k\cdot;n}, \hat{u}\right\rangle.$$

Proof of Lemma 6. The only interesting cases are $k=5$ with $\mu_{(5)}(t/n,a,\eta) = (t/n-\eta)_+^3$ and $k=6$ with $\mu_{(6)}(t/n,a,\eta) = -a_5 3(t/n-\eta)_+^2$, for $k=1,\ldots,4$ we have simply $\mathbf{r}_{k\cdot;n}=0$.

For $k=5$ and $t/n \ge \eta+\delta$, we have

$$\mu_{(5)}(t/n,\theta+\hat{u}) - \mu_{(5)}(t/n,\theta) = -3\hat{u}_6(t/n-\eta)^2 + 3\hat{u}_6^2(t/n-\eta) - \hat{u}_6^3.$$

For $k=5$ and $t/n < \eta+\delta$, we obtain by the mean value theorem

$$\mu_{(5)}(t/n,\theta+\hat{u}) - \mu_{(5)}(t/n,\theta) = -3(t/n-\eta-u_6(t))_+^2\hat{u}_6$$

where $|u_6(t)| \le |\hat{u}_6|$. Hence $\mathbf{r}_{5i;n}=0$ for all $i=1,\ldots,5$ and

$$\mathbf{r}_{56;n} = -3\sum_{t=n(\delta+\eta)}^{n} \xi_t\left[(t/n-\eta)^2 - \hat{u}_6(t/n-\eta) + \hat{u}_6^2\right] - 3\sum_{t\in nB(\eta,\delta)} \xi_t(t/n-\eta-u_6(t))_+^2.$$

For $k=6$ and $t/n > \eta+\delta$, we have

$$\begin{aligned}\mu_{(6)}(t/n,\theta+\hat{u}) - \mu_{(6)}(t/n,\theta) &= -(a_5+\hat{u}_5)3(t/n-(\eta+\hat{u}_6))^2 + a_5 3(t/n-\eta)^2\\ &= -3\,\hat{u}_5\left(t/n-\eta-\hat{u}_6\right)^2 - 3\,a_5\left[(t/n-\eta-\hat{u}_6)^2 - (t/n-\eta)^2\right]\\ &= -3\,\hat{u}_5\left[(t/n-\eta)^2 - 2\,\hat{u}_6(t/n-\eta) + \hat{u}_6^2\right] - 3\,a_5\left[-2\,\hat{u}_6(t/n-\eta) + \hat{u}_6^2\right].\end{aligned}$$

If $t/n < \eta+\delta$ then

$$\begin{aligned}\mu_{(6)}(t/n,\theta+\hat{u}) - \mu_{(6)}(t/n,\theta) &= -3\,\hat{u}_5\left(t/n-\eta-\hat{u}_6\right)_+^2 - 3\,a_5\left[(t/n-\eta-\hat{u}_6)_+^2 - (t/n-\eta)_+^2\right]\\ &= -3\,\hat{u}_5\left(t/n-\eta-\hat{u}_6\right)_+^2 + 6\,a_5\hat{u}_6(t/n-\eta-u_6^*(t))_+\end{aligned}$$

where $|u_6^*(t)| \le |\hat{u}_6|$. Hence $\mathbf{r}_{6i;n}=0$ for all $i=1,\ldots,4$ and

$$\mathbf{r}_{65;n} = -3\sum_{t\ge n(\delta+\eta)} \xi_t\left[(t/n-\eta)^2 - 2\,\hat{u}_6(t/n-\eta) + \hat{u}_6^2\right] - 3\sum_{\frac{t}{n}\in B(\eta,\delta)} \xi_t\left(t/n-\eta-\hat{u}_6\right)_+^2,$$

$$\mathbf{r}_{66;n} = -3\,a_5\sum_{t\ge n(\delta+\eta)} \xi_t\left[\hat{u}_6 - 2(t/n-\eta)\right] + 6\,a_5\sum_{\frac{t}{n}\in B(\eta,\delta)} \xi_t(t/n-\eta-u_6^*(t))_+. \quad \square$$

Lemma 7. *Define the matrix* $\mathbf{R_n} := \left[\mathbf{r_{kl;n}}\right]_{k=1,\ldots,6;l=1,\ldots,6} \in \mathbb{R}^{6\times 6}$. *Then*

$$\mathbb{P}\left(n^{-1}\|\mathbf{R_n}\| > \Delta\right) = O(\gamma^{\kappa}(n))$$

where $\kappa = \min(j,m)$ *and* $\|\mathbf{R_n}\| = \sqrt{\lambda_{\max}}$ *and* $\lambda_{\max}$ *is the largest eigenvalue of* $\mathbf{R_n'}\mathbf{R_n}$.

Proof of Lemma 7. Let $\Delta > 0$ be fixed. It sufficies to show that

$$\mathbb{P}\left(n^{-1}\left|\mathbf{r}_{ij;n}\right| > 2\Delta\right) = O\left(\gamma^{\kappa}(n)\right)$$

for $i,j=5,6$. Let $K := 3\cdot\max(2, 2|a_5|)$ and let $U_0 := \left\{\|\hat{u}\| < \delta\right\}$ for $\delta < 1$ such that $1+\mathrm{Var}(\xi_t) < \frac{\Delta^2}{16\delta^2K^2}$. For $0<\varepsilon<\frac{\Delta}{3K}$ define the sets

$$A_k := \left\{n^{-1}\left|\sum_{t>n(\eta+\delta)} \xi_t\,(t/n-\eta)^k\right| < \varepsilon\right\}, \quad k=0,1,2$$

and define $V_0 := \left\{2\lceil(n+1)\delta\rceil^{-1}\sum_{t\in nB(\eta,\delta)}\xi_t^2 \le \mathrm{Var}(\xi_t)+1\right\}$. If we assume that $\omega \in A_0 \cap A_1 \cap A_2 \cap V_0 \cap U_0$, we obtain

$$\begin{aligned} n^{-1}\left|\mathbf{r}_{ij;n}(\omega)\right| &\le n^{-1}\max\left(1, \left|\hat{u}_6(\omega)\right|\right) K \sum_{k=0}^{2}\left|\sum_{t\ge n(\delta+\eta)} \xi_t(\omega)\,(t/n-\eta)^k\right| + n^{-1}K\sqrt{\sum_{\frac{t}{n}\in B(\eta,\delta)} 1 \sum_{\frac{t}{n}\in B(\eta,\delta)} \xi_t^2(\omega)} \\ &\le 3K\varepsilon + n^{-1}K2\lceil(n+1)\delta\rceil\sqrt{\mathrm{Var}(\xi_t)+1} \\ &< \Delta + \frac{\lceil(n+1)\delta\rceil}{n\delta}\frac{\Delta}{2} < 2\Delta \quad \text{for } n \text{ large,} \end{aligned}$$

i.e.

$$A_0 \cap A_1 \cap A_2 \cap V_0 \cap U_0 \subset \{n^{-1}|\mathbf{r}_{ij;n}| \le 2\Delta\}$$

for n large. By Lemma 5 and Chebyshev's inequality, we have $\mathbb{P}(A_i^c) = O(\gamma^m(n))$ for $k = 0, \ldots, 2$. Theorem 1 implies $\mathbb{P}(U_0^c) = O(\gamma^\kappa(n))$ and again by Chebyshev's inequality $\mathbb{P}(V_0^c) = O(\gamma^j(n))$. Now, for any $\Delta > 0$ we can choose $\varepsilon > 0$ and $\delta \in (0,1)$ such that and so

$$\mathbb{P}\left(n^{-1}\left|\mathbf{r}_{ij;n}\right| > 2\Delta\right) = O(\gamma^\kappa(n)). \quad \square$$

Lemma 8. *There exists a constant depending only on a and η such that*

$$\mathbb{P}\left(\left\|\hat{u}\right\| > Cn^{-1}\left\|\mathbf{M}_\mathbf{n}'\mathbf{e}_\mathbf{n}\right\|\right) = O(\gamma^\kappa(n))$$

where $\kappa = \min(j, m)$.

Proof of Lemma 8. We obtain $\hat{u}$ by minimizing $L(\tau)$, i.e. $\nabla L(\theta+\hat{u}) = 0$. Hence

$$\begin{aligned} 0 = \frac{\partial}{\partial\theta_k}L(\theta+\hat{u}) &= 2\sum_{t=1}^{n}\left[X_t - \mu(t/n,\theta+\hat{u})\right]\mu_{(k)}(t/n,\theta+\hat{u}) \\ &= 2\sum_{t=1}^{n}\xi_t\mu_{(k)}(t/n,\theta+\hat{u}) + 2\sum_{t=1}^{n}\mu_{(k)}(t/n,\theta+\hat{u})\left[\mu(t/n,\theta) - \mu(t/n,\theta+\hat{u})\right] \\ &= 2\sum_{t=1}^{n}\xi_t\mu_{(k)}(t/n,\theta) + 2\sum_{t=1}^{n}\xi_t\left[\mu_{(k)}(t/n,\theta+\hat{u}) - \mu_{(k)}(t/n,\theta)\right] - 2\sum_{t=1}^{n}\mu_{(k)}(t/n,\theta+\hat{u})\nabla\mu(t/n,\theta+\tilde{u}(t))'\hat{u}, \end{aligned}$$

i.e.

$$\mathbf{M}_\mathbf{n}'\mathbf{e}_\mathbf{n} = -\mathbf{R}_\mathbf{n}\hat{u} + \hat{\mathbf{M}}_\mathbf{n}'\tilde{\mathbf{M}}_\mathbf{n}\hat{u} = \left[\mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n} - \mathbf{R}_\mathbf{n} + (\hat{\mathbf{M}}_\mathbf{n}'\tilde{\mathbf{M}}_\mathbf{n} - \mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n})\right]\hat{u} \tag{18}$$

where $\hat{\mathbf{M}}_\mathbf{n} := \left(\mu_{(k)}(t/n,\theta+\hat{u})\right)_{t,k}$, $\tilde{\mathbf{M}}_\mathbf{n} := \left(\mu_{(k)}(t/n,\theta+\tilde{u}(t))\right)_{t,k}$. Since $\frac{1}{n}\mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n}$ has full rank, $n^{-1}[\mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n} - \mathbf{R}_\mathbf{n} + (\hat{\mathbf{M}}_\mathbf{n}'\tilde{\mathbf{M}}_\mathbf{n} - \mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n})]$ is invertible with $\left\|\left[n^{-1}\mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n} - n^{-1}\mathbf{R}_\mathbf{n} + n^{-1}(\hat{\mathbf{M}}_\mathbf{n}'\tilde{\mathbf{M}}_\mathbf{n} - \mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n})\right]^{-1}\right\| \le C$, if both $n^{-1}\|\mathbf{R}_\mathbf{n}\|$ and $n^{-1}\|\mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n} - \hat{\mathbf{M}}_\mathbf{n}'\tilde{\mathbf{M}}_\mathbf{n}\|$ are sufficiently small, say for instance less than Δ^*. In this case

$$\left\|\hat{u}\right\| \le n^{-1}\|\mathbf{M}_\mathbf{n}\mathbf{e}_\mathbf{n}\|\, C$$

and so

$$\mathbb{P}\left(\left\|\hat{u}\right\| > Cn^{-1}\left\|\mathbf{M}_\mathbf{n}'\mathbf{e}_\mathbf{n}\right\|\right) \le \mathbb{P}\left(n^{-1}\|\mathbf{R}_\mathbf{n}\| > \Delta^*\right) + \mathbb{P}\left(n^{-1}\left\|\mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n} - \hat{\mathbf{M}}_\mathbf{n}'\tilde{\mathbf{M}}_\mathbf{n}\right\| > \Delta^*\right).$$

Applying Lemmas 7 and 9 concludes the proof. $\square$

Lemma 9.

$$\mathbb{P}\left(n^{-1}\left\|\mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n} - \hat{\mathbf{M}}_\mathbf{n}'\tilde{\mathbf{M}}_\mathbf{n}\right\| > \Delta\right) = O(\gamma^\kappa(n))$$

where $\kappa = \min(j, m)$.

Proof of Lemma 9. It is sufficient to show that

$$\mathbb{P}\left(n^{-1}\left[\mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n} - \hat{\mathbf{M}}_\mathbf{n}'\tilde{\mathbf{M}}_\mathbf{n}\right]_{j;k} > \Delta^*\right) = O(\gamma^\kappa(n)) \quad \forall j, k = 1, \ldots, 6.$$

However, for any compact $K \subset \mathbb{R}^6$, the gradient $\nabla_\theta \mu(\cdot,\cdot) : [0,1] \times K \to \mathbb{R}^6$ is uniformly continuous. Hence

$$\mathbb{P}\left(n^{-1}\left[\mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n} - \hat{\mathbf{M}}_\mathbf{n}'\tilde{\mathbf{M}}_\mathbf{n}\right]_{j;k} > \Delta^*\right) \le \mathbb{P}\left(\|\hat{u}\| > \delta^*\right) = O(\gamma^\kappa(n)). \quad \square$$

Proof of Theorem 2. Define the event

$$B_0 := \left\{\left\|\hat{u}\right\| > Cn^{-1}\left\|\mathbf{M}_\mathbf{n}'\mathbf{e}_\mathbf{n}\right\|\right\}.$$

By Eq. (18)

$$\mathbb{P}\left(\gamma^{-\frac{m}{2}}(n)\left|\hat{u} - (\mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n})^{-1}\mathbf{M}_\mathbf{n}'\mathbf{e}_\mathbf{n}\right| > 2\Delta\right) \le \mathbb{P}\left(\gamma^{-\frac{m}{2}}(n)\ \left\|(\mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n})^{-1}\right\|\ \|\mathbf{R}_\mathbf{n}\|\ \|\hat{u}\| > \Delta\right)$$
$$+\mathbb{P}\left(\gamma^{-\frac{m}{2}}(n)\ \left\|(\mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n})^{-1}\right\|\ \|\hat{\mathbf{M}}_\mathbf{n}'\tilde{\mathbf{M}}_\mathbf{n} - \mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n}\|\ \|\hat{u}\| > \Delta\right) = \Pi_1 + \Pi_2.$$

However

$$\Pi_1 \le \mathbb{P}(B_0) + \mathbb{P}\left(\left\{\gamma^{-\frac{m}{2}}(n)\ \left\|(\mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n})^{-1}\right\|\ \|\mathbf{R}_\mathbf{n}\|\ \|\hat{u}\| > \Delta\right\} \cap B_0^c\right)$$
$$\le \mathbb{P}(B_0) + \mathbb{P}\left(C\left\|(n^{-1}\mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n})^{-1}\right\|\ n^{-1}\|\mathbf{R}_\mathbf{n}\|\ \gamma^{-\frac{m}{2}}(n)\, n^{-1}\left\|\mathbf{M}_\mathbf{n}'\mathbf{e}_\mathbf{n}\right\| > \Delta\right).$$

By Lemma 5 $\left(\gamma^{\frac{m}{2}}(n)n^{-1}\sum_{t=1}^n \mu_{(k)}(t/n,\theta)\xi_t\right)_{n\in\mathbb{N}} = O_p(1)$, i.e.

$$\gamma^{\frac{m}{2}}(n)n^{-1}\left\|\mathbf{M}_\mathbf{n}'\mathbf{e}_\mathbf{n}\right\| = O_p(1).$$

By Lemma 7 $n^{-1}\|\mathbf{R}_\mathbf{n}\| = o_p(1)$ and so $\Pi_1 = o(1)$. We can estimate Π_2 in the same way, if we apply Lemma 9 instead of Lemma 7. $\square$

Proof of Theorem 3. Let Σ_n denote the covariance matrix of $\mathbf{e}_\mathbf{n}$. Then

$$\operatorname{Cov}\left(\gamma^{-\frac{m}{2}}(n)(\mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n})^{-1}\mathbf{M}_\mathbf{n}'\mathbf{e}_n\right) = \gamma^{-m}(n)n(\mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n})^{-1}\frac{1}{n^2}\mathbf{M}_\mathbf{n}'\Sigma\mathbf{M}_\mathbf{n}n(\mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n})^{-1}$$
$$= \Lambda(1+o(1))\frac{\gamma^{-m}(n)}{n^2}\mathbf{M}_\mathbf{n}'\Sigma_n\mathbf{M}_\mathbf{n}\Lambda(1+o(1)).$$

Expanding the middle expression yields

$$\frac{\gamma^{-m}(n)}{n^2}\left(\mathbf{M}_\mathbf{n}'\Sigma\mathbf{M}_\mathbf{n}\right)_{jk} = \frac{\gamma^{-m}(n)}{n^2}\sum_{s,t=1}^n \mu_{(j)}(t/n)\sigma_{t,s}\mu_{(k)}(s/n)$$
$$= \frac{\gamma^{-m}(n)}{n^2}\sum_{s,t=1}^n\sum_{l=m}^\infty \frac{C_l^2}{l!}\gamma^l(s-t)\mu_{(k)}(s/n)\mu_{(j)}(t/n)$$
$$= \frac{\gamma^{-m}(n)}{n^2}\frac{C_m^2}{m!}\sum_{s,t=1}^n \gamma^m(s-t)\mu_{(k)}(s/n)\mu_{(j)}(t/n)$$
$$+\sum_{l=m+1}^\infty \frac{\gamma^{-m}(n)}{n^2}\sum_{s,t=1}^n \frac{C_l^2}{l!}\gamma^l(s-t)\mu_{(k)}(s/n)\mu_{(j)}(t/n)$$
$$= S_{jk}^m + S_{jk}^0.$$

By Lemma 5

$$\lim_n S_{jk}^m = \frac{C_m^2}{m!}\lim_n\left(\frac{\gamma^{-m}(n)}{n^2}\sum_{s,t=1}^n \gamma^m(s-t)\mu_{(k)}(s/n)\mu_{(j)}(t/n)\right)$$
$$= \frac{C_m^2}{m!}\int_0^1 \frac{\mu_{(j)}(t)\mu_{(k)}(s)}{|s-t|^{\alpha m}}\,\mathrm{d}t\,\mathrm{d}s.$$

For estimating the term S_{jk}^0, we note that for all $\varepsilon > 0$ there is a $n_\varepsilon > 0$ such that $|\gamma(n)| < \varepsilon$ for all $n > n_\varepsilon$. Hence

$$S_{jk}^0 = \sum_{l=m+1}^\infty \frac{\gamma^{-m}(n)}{n^2}\sum_{s,t=1,|s-t|>n_\varepsilon}\frac{C_l^2}{l!}\gamma^l(s-t)\mu_{(k)}(s/n)\mu_{(j)}(t/n)$$

$$
\begin{aligned}
&+\sum_{l=m+1}^{\infty}\frac{\left|\gamma^{-m}(n)\right|}{n^2}\sum_{s,t=1,|s-t|\le n_\varepsilon}^{n}\frac{C_l^2}{l!}\gamma^l(s-t)\mu_{(k)}(s/n)\mu_j(t/n)\\
&\le \sum_{l=m+1}^{\infty}\frac{\left|\gamma^{-m}(n)\right|}{n^2}\varepsilon\sum_{s,t=1}^{n}\frac{C_l^2}{l!}\left|\gamma^m(s-t)\right|\left|\mu_{(k)}(s/n)\mu_{(j)}(t/n)\right|\\
&\quad+\gamma^m(0)\sum_{l=m+1}^{\infty}\frac{\left|\gamma^{-m}(n)\right|}{n^2}\sum_{s,t=1,|s-t|\le n_\varepsilon}^{n}\frac{C_l^2}{l!}\left|\mu_{(k)}(s/n)\mu_{(j)}(t/n)\right|\\
&= A_n+B_n.
\end{aligned}
$$

An upper bound for the first term is given by

$$
A_n \le K\varepsilon\sum_{s,t=1}^{n}\left|L(|s-t|)\frac{\gamma^{-m}(n)}{n^2}\gamma^m(s-t)\mu_{(k)}(s/n)\mu_{(j)}(t/n)\right| = O(\varepsilon).
$$

For the second term, note that n_ε is fixed,

$$
n_\varepsilon \lim_{n\to\infty}\sum_{u=1}^{n}\mu_{(j)}\left(\frac{u}{n}\right)n^{-1} \to n_\varepsilon\int_0^1 \mu_{(j)}(t)\,\mathrm{d}t,
$$

and

$$
K=\sum_{l=m+1}^{\infty}\frac{C_l^2}{l!}<\infty.
$$

Then

$$
\begin{aligned}
B_n &\le \frac{\gamma^{-m}(n)}{n}K\sum_{u=1}^{n}\sum_{v=\max(1,u-n_\varepsilon)}^{\min(n,u+n_\varepsilon)}\mu_{(j)}(u/n)\mu_{(k)}(v/n)\frac{1}{n}\\
&\le \frac{\gamma^{-m}(n)}{n}K\cdot 2n_\varepsilon\sum_{u=1}^{n}\mu_{(j)}\left(\frac{u}{n}\right)n^{-1}.
\end{aligned}
$$

Since $\alpha > m^{-1}$ we have $n\gamma^m(n)\to 0$ so that $B_n = o(1)$. Thus S_{jk}^0 converges to zero which proves (9). Furthermore

$$
\begin{aligned}
&\mathbb{E}(\gamma^{-\frac{m}{2}}(n)(\mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n})^{-1}\mathbf{M}_\mathbf{n}'\mathbf{e}_n)=0,\\
&\gamma^{-m}(n)\mathrm{Cov}\left((\mathbf{M}_\mathbf{n}'\mathbf{M}_\mathbf{n})^{-1}\mathbf{M}_\mathbf{n}'\mathbf{e}_n - Y_n\right)=S_{jk}^0\to 0,
\end{aligned}
$$

from which (8) follows directly. □

Proof of Corollary 2. Define $S_n(s) := \sum_{v=0}^{\lfloor ns\rfloor}\xi_n$. By construction, $S_n(s)$ is a stochastic process with piecewise constant sample paths and jumps at $s=t/n$ with size ξ_t. Hence, we can write

$$
\frac{\gamma^{-\frac{m}{2}}(n)}{n}\left(\mathbf{M}_\mathbf{n}'\mathbf{e}_\mathbf{n}\right)_k = \frac{\gamma^{-\frac{m}{2}}(n)}{n}\sum_{t=1}^{n}\mu_{(k)}(t/n)\xi_t = \frac{\gamma^{-\frac{m}{2}}(n)}{n}\int_0^1\mu_{(k)}(s)\,\mathrm{d}S_n(s).
$$

Applying integration by parts leads to

$$
\begin{aligned}
\frac{\gamma^{-\frac{m}{2}}(n)}{n}\int_0^1\mu_{(k)}(s)\,\mathrm{d}S_n(s) &= \frac{\gamma^{-\frac{m}{2}}(n)}{n}\left(-\int_0^1 S_n(s)\,\mathrm{d}\mu_{(k)}(s)+S_n(1)\mu_{(k)}(1)\right)\\
&= \left(-\int_0^1\frac{\gamma^{-\frac{m}{2}}(n)}{n}S_n(s)\,\mathrm{d}\mu_{(k)}(s)+\frac{\gamma^{-\frac{m}{2}}(n)}{n}S_n(1)\mu_{(k)}(1)\right)\\
&=: h_k\left(\frac{\gamma^{-\frac{m}{2}}(n)}{n}S_n(s)\right).
\end{aligned}
$$

Theorem 5.6 in [51] states that

$$
\frac{\gamma^{-\frac{m}{2}}(n)}{n}S_n(\cdot)\Longrightarrow\frac{C(m)}{m!}\mathcal{H}_m(\cdot)
$$

where "$\Longrightarrow$" denotes weak convergence in the space $D([0,1])$ endowed with the Skorohod topology. Lemmas 10 and 11 below imply that the functional

$$h_{(k)}: D([0,1]) \longrightarrow \mathbb{R}: f \longmapsto -\int f(s)\,\mathrm{d}\mu_{(k)}(s) + \mu_{(k)}(1)f(1)$$

is continuous. Hence, by the continuous mapping theorem

$$h_{(k)}\left(\frac{\gamma^{-\frac{m}{2}}(n)}{n}S_n(\cdot)\right) \to_d h_{(k)}\left(\frac{C(m)}{m!}\mathcal{H}_m(\cdot)\right). \quad \square \tag{19}$$

Lemma 10. *If $f_n \longrightarrow f$ within $D([0,1])$, then $f_n(1) \longrightarrow f(1)$.*

Proof of Lemma 10. Follows from the discussion in [10, page 121]. □

Lemma 11. *The mapping $m_{(k)}: f \longmapsto \int_0^1 f(s)\,\mathrm{d}\mu_{(k)}(s)$ is continuous.*

Proof of Lemma 11. Let $f_n \to f$ in $D([0,1])$. There exists a sequence $(\lambda_n)_{n\in\mathbb{N}}$ of monotonous, continuous bijective functions on $[0,1]$ such that $\lambda_n \to \mathrm{id}$ uniformly and $f_n \circ \lambda_n \to f$ uniformly, c.f. [10, page 111]. Since $s \mapsto \mu_{(k)}(s)$ is absolutely continuous, we obtain

$$\begin{aligned} |m_{(k)}(f) - m_{(k)}(f_n)| &= \left|\int_0^1 (f(t) - f_n(t))\frac{\partial}{\partial t}\mu_{(k)}(t)\,\mathrm{d}t\right| \\ &\le \left|\int_0^1 (f(t) - f(\lambda_n^{-1}(t)))\frac{\partial}{\partial t}\mu_{(k)}(t)\,\mathrm{d}t\right| + \left|\int_0^1 (f(\lambda_n^{-1}(t)) - f_n(t))\frac{\partial}{\partial t}\mu_{(k)}(t)\,\mathrm{d}t\right| \\ &= A + B, \end{aligned}$$

because $f \in D([0,1])$, f is bounded and has only countable but many discontinuities; see [10, page 110]. Since

$$\sup_{x\in[0,1]}\left|\lambda_n^{-1}(x) - x\right| = \sup_{x\in[0,1]}\left|\lambda_n^{-1}(\lambda_n(x)) - \lambda_n(x)\right| = \sup_{x\in[0,1]}|x - \lambda_n(x)| \to 0,$$

λ_n^{-1} converges uniformly to the identity function as well. We can thus apply Lebesgue's theorem and conclude that $A \to 0$. Furthermore, we obtain the upper bound

$$\begin{aligned} B &\le \left\|\frac{\partial}{\partial t}\mu_{(k)}\right\|_1 \left\|f \circ \lambda_n^{-1} - f_n\right\|_\infty = \left\|\frac{\partial}{\partial t}\mu_{(k)}\right\|_1 \sup_{x\in[0,1]}\left|f(\lambda_n^{-1}(x)) - f_n(x)\right| \\ &= \left\|\frac{\partial}{\partial t}\mu_{(k)}\right\|_1 \sup_{x\in[0,1]}|f(x) - f_n(\lambda_n(x))| \to 0. \quad \square \end{aligned}$$

A.3. Modifications for quadratic splines

The proofs given above carry over to quadratic splines, with the following modifications. The regression function is given by

$$\mu(s,\theta) = a_1 + a_2 s + a_3 s^2 + a_4(s-\eta)_+^2.$$

In particular, the derivatives with respect to a_4 and η are equal to

$$\mu_{(4)}(s,\theta+u) = (s - \eta - u_5)_+^2$$

and

$$\mu_{(5)}(s,\theta+u) = 2(a_4+u_4)(s-\eta-u_5)_+$$

respectively. Lemma 6 is replaced by

Lemma 6b. *For all $k = 1,\ldots,5$ there exists a vector $\mathbf{r}_{k\cdot;n} \in \mathbb{R}^6$ such that*

$$\sum_{t=1}^n \xi_t\left[\mu_{(k)}(t/n,\theta+\hat{u}) - \mu_{(k)}(t/n,\theta)\right] = \langle \mathbf{r}_{k\cdot;n}, \hat{u}\rangle.$$

Proof of Lemma 6b. The case $k = 1, \ldots, 3$ is again clear. The case $k = 4$ is similar to the case $k = 5$ for the cubic spline. It remains to proof the case $k = 5$ for the quadratic spline. The result can be obtained explicitly as follows:

$$\begin{aligned}
&\sum_{t=1}^{n} \xi_t \left[2(a_4 + \hat{u}_4)(t/n - \eta - \hat{u}_5)_+ - 2a_4(t/n - \eta)_+\right] \\
&= 2\sum_{t=1}^{n} \xi_t \left[\hat{u}_4(t/n - \eta - \hat{u}_5)_+\right] + \sum_{t=1}^{n} \xi_t \left[2a_4 \left((t/n - \eta - \hat{u}_5)_+ - (t/n - \eta)_+\right)\right] \\
&= 2\hat{u}_4 \left[\sum_{t/n \geq \eta+\delta} \xi_t(t/n - \eta) + \sum_{t/n \geq \eta+\delta} \xi_t \hat{u}_5 \sum_{t \in nB(\eta,\delta)} \xi_t(t/n - \eta - \hat{u}_5)_+\right] \\
&\quad + 2\hat{u}_5 a_4 \left[\sum_{t/n \geq \eta+\delta} \xi_t + \sum_{t \in nB(\eta,\delta)} \frac{(t/n - \eta - \hat{u}_5)_+ - (t/n - \eta)_+}{\hat{u}_5}\right] \\
&= \langle \mathbf{r}_{5\cdot;n}, \hat{u} \rangle
\end{aligned}$$

with

$$\mathbf{r}_{5,4;n} = 2\left[\sum_{t/n \geq \eta+\delta} \xi_t(t/n - \eta) + \sum_{t/n \geq \eta+\delta} \xi_t \hat{u}_5 + \sum_{t \in nB(\eta,\delta)} \xi_t(t/n - \eta - \hat{u}_5)_+\right]$$

and

$$\mathbf{r}_{5,5;n} = 2a_4 \left[\sum_{t/n \geq \eta+\delta} \xi_t + \sum_{t \in nB(\eta,\delta)} \xi_t \frac{(t/n - \eta - \hat{u}_5)_+ - (t/n - \eta)_+}{\hat{u}_5}\right]. \quad \square$$

The proof of Lemma 7 given above for cubic splines applies almost without changes to the quadratic case: We need to show that

$$\mathbb{P}\left(n^{-1} \left|\mathbf{r}_{5,j;n}\right| \geq \Delta\right) = O(\gamma^{\kappa}(n))$$

for $j = 4, 5$. Now, if $t \in nB(\eta, \delta)$ and $\hat{u} \leq \delta$ then $\left[(t/n - \eta - \hat{u}_5)_+ - (t/n - \eta)_+\right]/\hat{u}_5$ and $(t/n - \eta - \hat{u}_5)_+$ are bounded, so that the same arguments as in the case of cubic splines can be applied.

References

[1] J. Beran, Statistics for Long-Memory Processes, Chapman and Hall, London, 1994.
[2] J. Beran, M-estimators of location for data with slowly decaying serial correlations, J. Amer. Statist. Assoc. 86 (1991) 704–708.
[3] J. Beran, R.J. Bhansali, D. Ocker, On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes, Biometrika 85 (1998) 921–934.
[4] J. Beran, Y. Feng, Local polynomial estimation with a FARIMA-GARCH error process, Bernoulli 7 (5) (2001) 733–750.
[5] J. Beran, Y. Feng, SEMIFAR models—a semiparametric framework for modelling trends, long-range dependence and nonstationarity, Comput. Statist. Data Anal. 40 (2) (2002) 393–419.
[6] J. Beran, Y. Feng, Data driven bandwidth choice for SEMIFAR models, J. Comput. Graph. Statist. 11 (2) (2002) 690–713.
[7] J. Beran, Y. Feng, Local polynomial fitting with long memory, short memory andantipersistent errors, Ann. Inst. Statist. Math. 54 (2) (2002) 291–311.
[8] J. Beran, Y. Feng, Weighted averages and local polynomial estimation for fractional linear ARCH processes, J. Stat. Theory Pract. 1 (2) (2007) 149–166.
[9] J. Beran, S. Ghosh, Root-n-consistent estimation in partial linear models with long-memory errors, Scand. J. Statist. 25 (1998) 345–357.
[10] P. Billingsley, Convergence of Probability Measures, Wiley, New York, 1968.
[11] P. Billingsley, F. Topsoe, Uniformity in weak convergence, Probab. Theory Related Fields 7 (1) (1967) 1–16.
[12] N.H. Bingham, C.M. Goldie, J.L. Teugels, Regular Variation, Cambridge University Press, Cambridge, 1987.
[13] P. Breuer, P. Major, Central limit theorems for non-linear functionals of Gaussian fields, J. Multivariate Anal. 13 (3) (1983) 425–441.
[14] B.E. Brodsky, B.S. Darkhosky, Non-Parametric Statistical Diagnosis: Problems and Methods, Springer, New York, 2000.
[15] J. Chen, A.K. Gupta, Parametric Statistical Change Point Analysis (Oberwolfach Seminars), Birkhäuser, Basel, 2000.
[16] S. Csörgö, L. Horvath, Limit Theorems in Change-Point Analysis, Wiley, New York, 1998.
[17] S. Csörgö, J. Mielniczuk, Nonparametric regression under long-range dependent normal errors, Ann. Statist. 23 (3) (1995) 1000–1014.
[18] S. Csörgö, J. Mielniczuk, Random-design regression under long-range dependent errors, Bernoulli 5 (2) (1999) 209–224.
[19] R. Dahlhaus, Efficient location and regression estimation for long range dependent regression models, Ann. Statist. 23 (1995) 1029–1047.
[20] C. de Boor, A Practical Guide to Splines, Springer, New York, 2001.
[21] P.J. Diggle, M.F. Hutchinson, On spline smoothing with autocorrelated errors, Austral. J. Statist. 31 (1989) 166–182.
[22] R.L. Dobrushin, P. Major, Non-central limit theorems for non-linear functionals of Gaussian fields, J. Probab. Theory Related Fields 50 (1979) 27–52.
[23] P. Doukhan, G. Oppenheim, M.S. Taqqu, Theory and Application of Long-Range Dependence, Birkhäuser, Boston, 2002.
[24] R.L. Eubank, Nonparametric Regression and Spline Smoothing, second ed., Marcel Dekker, New York, 1999.
[25] P.I. Feder, On asymptotic distribution theory in segmented regression problems—identified case, Ann. Statist. 3 (1) (1975) 49–83.
[26] R.F. Galán, M. Weidert, R. Menzel, A.V.M. Herz, C.G. Galizia, Sensory memory for odors is encoded in spontaneous correlated activity between olfactory Glomeruli, Neural Comput. 18 (2006) 10–25.
[27] C.G. Galizia, R. Menzel, The role of Glomeruli in the neural representation of odours: results from optical recording studies, J. Insect Physiol. 47 (2001) 115–130.
[28] A.R. Gallant, Inference for nonlinear models, in: Mimeograph Series No. 875, Institute of Statistics, North Carolina State University, Raleigh, 1973.

[29] A.R. Gallant, The theory of nonlinear regression as it relates to segmented polynomial regression with estimated join points, in: Mimeograph Series No. 925, Institute of Statistics, North Carolina State University, Raleigh, 1974.
[30] J.T. Gao, V.V. Anh, Semiparametric regression under long-range dependent errors, J. Statist. Plann. Inference 80 (1999) 37–57.
[31] C.W.J. Granger, R. Joyeux, An introduction to long-memory time series, J. Time Ser. Anal. 1 (1980) (1980) 15–30.
[32] P.J. Green, B.W. Silverman, Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach, Chapman and Hall, London, 1994.
[33] P. Hall, J.D. Hart, Nonparametric regression with long-range dependence, Stochastic Process. Appl. 36 (1990) 339–351.
[34] J.R.M. Hosking, Fractional differencing, Biometrika 68 (1981) 165–176.
[35] A.V. Ivanov, N.N. Leonenko, Asymptotic inference for a nonlinear regression with long range dependent errors, Theory Probab. Math. Statist. 63 (2001) 65–83.
[36] A.V. Ivanov, N.N. Leonenko, Asymptotic theory of non-linear regression with long range dependent errors, Math. Methods Statist. 13 (2) (2004) 153–178.
[37] A.V. Ivanov, N.N. Leonenko, Semiparametric analysis of long range dependence in non-linear regression, J. Statist. Plann. Inference 138 (6) (2008) 1733–1753.
[38] J. Joerges, A. Küttner, C.G. Galizia, R. Menzel, Representations of odours and odour mixtures visualized in the honeybee brain, Nature 387 (1997) 285–288.
[39] J. Kim, H.J. Kim, Asymptotic results in segmented multiple regression, J. Multivariate Anal. 99 (2008) 2016–2038.
[40] R. Kohn, C.F. Ansley, C.M. Wong, Nonparametric spline regression with autoregressive moving average errors, Biometrika 79 (2) (1992) 335–346.
[41] H.L. Koul, Asymptotics of M-estimators in non-linear regression with long-range dependent errors, in: P.M. Robinson, M. Rosenblatt (Eds.), Athens Conference on Applied Probability and Time Series Volume II: Time Series Analysis in Memory of E.J. Hannan, in: Lecture Notes in Statistics, vol. 115, Springer, 1996, pp. 272–290.
[42] H.L. Koul, R.T. Baillie, Asymptotics of M-estimators in non-linear regression with long memory designs, Statist. Probab. Lett. 61 (2003) 237–252.
[43] H.R. Künsch, J. Beran, F. Hampel, Contrasts under long-range correlations, Ann. Statist. 21 (2) (1993) 943–964.
[44] J. Liu, S. Wu, J.V. Zidek, On segmented multivariate regression, Statist. Sinica 7 (1997) 497–525.
[45] S.B. Lowen, M.C. Teich, Fractal Based Point Processes, Wiley, New York, 2005.
[46] P. Major, Multiple Wiener-Ito integrals, in: Lecture Notes in Mathematics, vol. 849, Springer-Verlag, Berlin, 1981.
[47] J. Opsomer, Y. Wang, Y. Yang, Nonparametric regression with correlated errors, Statist. Sci. 16 (2) (2001) 134–153.
[48] P.M. Robinson, in: W. Barnett, J. Powell, G. Tauchen (Eds.), Nonparametric Function Estimation for Long-Memory Time Series. Nonparametric and Semiparametric Methods in Econometrics and Statistics, Cambridge University Press, 1991, pp. 437–457.
[49] G.A.F. Seber, C.J. Wild, Nonlinear Regression, Wiley, New York, 2003.
[50] M.S. Taqqu, Weak convergence to fractional Brownian motion and to the Rosenblatt process, J. Probab. Theory Related Fields 31 (4) (1975) 287–302.
[51] M.S. Taqqu, Convergence of integrated processes of arbitrary Hermite rank, J. Probab. Theory Related Fields 50 (1979) 53–83.
[52] G. Wahba, Improper priors, spline smoothingand the problem of guarding against model errors in regression, J. R. Stat. Soc. Ser. B 40 (1978) 364–372.
[53] G. Wahba, Spline Models for Observational Data, in: Regional Conference Series in Applied Mathematics, SIAM, 1990.
[54] Y. Wang, Smoothing splines models with correlated random errors, J. Amer. Statist. Assoc. 93 (441) (1998) 341–348.
[55] Y. Yajima, On estimation of a regression model with long term errors, Ann. Statist. 16 (2) (1988) 791–807.
[56] Y. Yajima, Asymptotic properties of the LSE in a regression model with long-memory stationary errors, Ann. Statist. 19 (1991) 158–177.