The Annals of Statistics
1995, Vol. 23, No. 3, 1029–1047

EFFICIENT LOCATION AND REGRESSION ESTIMATION FOR LONG RANGE DEPENDENT REGRESSION MODELS

BY RAINER DAHLHAUS

Universität Heidelberg

In this paper we construct an efficient weighted least squares estimator for the mean and more generally for the regression parameters in certain Gaussian long range dependent regression models, including polynomial regression. The form of the estimator does not depend on the whole dependence structure of the residuals, but only on the local behaviour of the spectral density at zero. By using an estimator of the self-similarity parameter, we give a fully efficient estimator. Furthermore, we construct efficient weighted M-estimators.

1. Introduction. Statistical models with long range dependent processes involved have recently been used in various fields, for example, in hydrology, economics or geophysics. Furthermore, they have turned out to be quite useful for modelling certain measurements that were originally supposed to be iid [cf. Hampel (1987)].

In this paper we consider the estimation of the mean of a long range dependent sequence and more generally of the coefficients in a regression model with long range dependent errors. Let ε_i, $i \in \mathbb{Z}$, be a stationary Gaussian process with mean zero, covariance function $c(u)$ and spectral density $f_\varepsilon(\lambda) = |\lambda|^{2\alpha} f_*(\lambda)$ with $\alpha \in (-1/2, \infty)$, where f_* is continuous in a neighborhood of 0 with $f_*(0) \neq 0$. If $\alpha \in (-1/2, 0)$, the covariances of the process are not summable and the process is called long range dependent. We assume that we observe

$$Y_i = X_i \beta + \varepsilon_i, \qquad i = 1, \dots, N,$$

and we want to estimate the parameter β. Let $\Sigma = \Sigma_N = \{\mathrm{cov}(\varepsilon_i, \varepsilon_j)\}_{i,j=1,\dots,N}$, $Y = (Y_1, \dots, Y_N)'$ and $X = (X_1', \dots, X_N')'$. If $X_i \equiv 1$, this is the problem of location estimation.

The problem of location and regression estimation has been studied by several authors before. Least squares estimators for linear regression models have been investigated by Yajima (1988, 1991), M-estimators by Koul (1992) and rank estimators and least absolute derivation estimators by Koul and Mukherjee (1993). Adenstedt (1974), Samarov and Taqqu (1988) and Beran and Künsch (1985) have pointed out that the arithmetic mean is no longer an efficient estimator for the location of a long range dependent process (which it is for a short range dependent process). As was pointed out by Yajima (1988),

Received January 1992; revised October 1994.
AMS 1991 *subject classifications.* Primary 62F12; secondary 60F99, 62M10
Key words and phrases. Long range dependence, weighted least squares estimates, M-estimates, efficiency.

the same holds for the case of polynomial regression with long range dependent errors: Again, the least square estimator is no longer efficient, while it is with short range dependent errors. While the efficiency loss for the arithmetic mean is minor, it is more dramatic for least squares polynomial regression.

The behaviour of least squares estimators for regression with stationary errors can best be seen from the following heuristics. The maximum likelihood estimator of β is

$$\bar{\beta}_N = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y \quad \text{with variance } (X'\Sigma^{-1}X)^{-1},$$

while the least squares estimator is

$$\beta_N^{\mathrm{LS}} = (X'X)^{-1}X'Y \quad \text{with variance } (X'X)^{-1}(X'\Sigma X)(X'X)^{-1}.$$

Suppose now that X is one dimensional and either stationary or asymptotically stationary with spectral measure $F_X(\lambda)$. (This assumption is only used in this heuristic.) Then [cf. Grenander (1954) and Yajima (1991)]

$$(X'X)^{-1}(X'\Sigma X)(X'X)^{-1} \sim \frac{2\pi}{N}\left(\int_{-\pi}^{\pi} dF_X(\lambda)\right)^{-2} \int_{-\pi}^{\pi} f_\varepsilon(\lambda)\, dF_X(\lambda)$$

and

$$(X'\Sigma^{-1}X)^{-1} \sim \frac{2\pi}{N}\left(\int_{-\pi}^{\pi} f_\varepsilon(\lambda)^{-1}\, dF_X(\lambda)\right)^{-1}.$$

Due to the Cauchy–Schwarz inequality, the former expression is larger or equal to the second and we can only achieve efficiency if $f_\varepsilon(\lambda) = \text{const } f_\varepsilon(\lambda)^{-1}$ a.s. with respect to the measure $F_X(\lambda)$, which includes two important special cases:

1. $f_\varepsilon(\lambda) = \text{constant}$, that is, the ε_i's are uncorrelated.
2. $F_X(\lambda)$ is a dirac measure at some frequency which includes the trigonometric design. $F_X(\lambda) = c\chi_{[0,\pi]}(\lambda)$ holds, for example, if $X_i = \varphi(i/N)$ with a smooth function φ which includes the location case $\varphi = 1$.

In the case where ε_i is long range dependent, $f_\varepsilon(0)$ no longer exists. Efficiency in the trigonometric case for frequencies different from zero still holds, which was proved by Yajima [(1991), Example 2.1]. However, in the case $F_X(\lambda) = c\chi_{[0,\pi]}(\lambda)$ we have to replace the "$\sim$" signs in the above heuristics by more refined asymptotics. As mentioned above, it is no longer true in this situation that the least squares estimator is efficient.

Adenstedt (1974) has shown that $\sum_{i=1}^{N} c_i Y_i$ with

$$c_i = \binom{N}{i} B(\alpha + i + 1, \alpha + N - i + 1)/B(\alpha + 1, \alpha + 1)$$

is an efficient estimator of the mean for any spectral density of the form $f_\varepsilon(\lambda) = |\lambda|^{2\alpha} f_*(\lambda)$. Since

$$\Gamma(x+\alpha)/\Gamma(x) \sim x^\alpha \quad \text{as } x \to \infty,$$

we obtain for $N \to \infty$, $i \to \infty$, $i/N \to x \in (0,1)$,

$$c_i \sim \text{const}\, w_\alpha(x),$$

where $w_\alpha(x) = x^\alpha(1-x)^\alpha$. This suggests to use

$$\left(\sum_{i=1}^N w_\alpha\left(\frac{i}{N+1}\right)\right)^{-1} \sum_{i=1}^N w_\alpha\left(\frac{i}{N+1}\right) Y_i$$

as a simplified estimate of the mean (Theorem 2.3 shows that it is efficient, too).

This estimator is a weighted least squares estimator, since it minimizes

$$\sum_{i=1}^N w_\alpha\left(\frac{i}{N+1}\right)(Y_i - X_i\,\beta)^2 \quad \text{for } X_i \equiv 1.$$

Therefore, the question arises for what other regressors X_i this weighted least squares estimator is still efficient. In this paper we show that this holds if the design matrix spans the same subspace as spanned by certain Jacobi polynomials. This includes the case of polynomial regressors.

In Section 2 we study the asymptotic behaviour of weighted least squares estimators. In particular, we prove efficiency in the situation mentioned above. It turns out that efficiency does not hold in all situations where $F_X(\lambda) = c\chi_{[0,\pi]}(\lambda)$. We give an example for this. Furthermore, we construct efficient weighted M-estimators.

If we estimate α from the data and replace α in w_α by the estimate $\hat{\alpha}$, we obtain a fully efficient estimator. This is proved in Section 3.

Large parts of the very technical proofs are put into the Appendix.

2. Efficiency of weighted least squares estimators.

Let

$$w,\ \varphi_j\colon [0,1] \to \mathbb{R},\ j = 1,\dots,p, \qquad X_i = X_{iN} = \big(\varphi_1(i/N),\dots,\varphi_p(i/N)\big),$$

$$I_w = \operatorname{diag}\left\{w\left(\frac{1}{N+1}\right),\dots,w\left(\frac{N}{N+1}\right)\right\}, \qquad Y_i = X_i\,\beta_0 + \varepsilon_i,$$

$$\begin{aligned}\hat{\beta}_N^{(w)} &= \arg\min(Y - X\beta)' I_w (Y - X\beta)\\ &= (X' I_w X)^{-1} X' I_w Y.\end{aligned}$$

Then

$$\begin{aligned}\operatorname{Cov}\big(\hat{\beta}_N^{(w)}\big) &= (X' I_w X)^{-1}(X' I_w \Sigma I_w X)(X' I_w X)^{-1}\\ &=: S_N^{-1} R_N S_N^{-1}.\end{aligned}$$

Furthermore, for the MLE $\bar{\beta}_N = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y$, we set

$$\operatorname{Cov}\big(\bar{\beta}_N\big) = (X'\Sigma^{-1}X)^{-1} =: \Gamma_N^{-1}.$$

We call an estimate efficient if it has asymptotically the same covariance matrix as the MLE. More sophisticated, it may be shown that the sequence of the local experiments is asymptotically normal (LAN) and that all estimators called efficient in this paper are locally asymptotically minimax [LAM; cf. Millar (1983)]. The same holds if α is no longer a fixed nuisance parameter, but estimated at the same time. This has been proved in Dahlhaus (1992).

In the following we therefore study the behaviour of $\operatorname{Cov}(\hat{\beta}_N^{(w)})$. To derive the asymptotic distribution of $\hat{\beta}_N^{(w)}$, we need the following definition. Let $\alpha \in (-1/2, \infty)$ and

$$k = k_\alpha = \begin{cases} [\alpha] + 1, & \alpha \notin \mathbb{N}, \\ \alpha, & \alpha \in \mathbb{N}_0. \end{cases}$$

A function $h: \mathbb{R} \to \mathbb{R}$ with $h(x) = 0$ for $x \notin [0, 1]$ is called of smoothness α if it fulfills the following conditions:

1. h is $(k-1)$-times continuously differentiable (in the case $k = 1$, we assume continuity, and in the case $k = 0$, we make no assumption) with $|h^{(k-1)}(x)| \le K \max\{|x|, 1 - |x|\}^{\alpha - k + 1}$.
2. There exists a finite set P such that h is $(k+1)$-times differentiable in all $x \notin P$ with $|h^{(k+1)}(x)| \le K \max\{|x|, 1 - |x|\}^{\alpha - k - 1}$.

Conditions 1 and 2 are, for example, fulfilled for $h(x) = x^\beta (1 - x)^\beta$ ($\beta \ge \alpha$; $P = \{0, 1\}$). Let $\Delta_{k,i}$ be the kth order difference operator applied to the argument i of a function, that is, $\Delta_{k,i}(f(i)) = \Delta_{k-1,i}(f(i)) - \Delta_{k-1,i}(f(i-1))$, $\Delta_{0,i}(f(i)) = f(i)$. A Taylor expansion leads, for a function h of smoothness α and $k = k_\alpha$, to

$$\Delta_{k,t}\left(h\left(\frac{t}{N+1}\right)\right) = O\left((N+1)^{-k} \sup_{[(t-k)/(N+1),\, t/(N+1)] \cap P^C} |h^{(k)}(x)|\right)$$

and, in the case $[(t-k)/(N+1), t/(N+1)] \cap P = \varnothing$,

$$\Delta_{k,t}\left(h\left(\frac{t}{N+1}\right)\right) = \frac{1}{(N+1)^k} h^{(k)}\left(\frac{t}{N+1}\right) + O\left((N+1)^{-k-1} \sup_{[(t-k)/(N+1),\, t/(N+1)]} |h^{(k+1)}(x)|\right).$$

THEOREM 2.1. *Let $f_\varepsilon(\lambda) \sim |\lambda|^{2\alpha}$ $(\lambda \to 0)$ with $\alpha \in (-1/2, \infty)$. Suppose $h_j(x) := \varphi_j(x) w(x)$ is of smoothness α. Then*

$$\hat{\beta}_N^{(w)} \sim_{\mathscr{D}} \mathscr{N}(\beta, S_N^{-1} \Sigma_N S_N^{-1})$$

with

$$S_{N_{ij}} \sim N \int_0^1 \varphi_i(x) \varphi_j(x) w(x)\, dx$$

and

$$\begin{aligned} R_{N_{ij}} &\sim N^{-2\alpha+1} 2 f_*(0) \Gamma(2\alpha - 2k + 1) \cos\left\{(2\alpha - 2k + 1)\frac{\pi}{2}\right\} \\ &\quad \times \int_0^1 \int_0^1 h_i^{(k)}(x) h_j^{(k)}(y) |x - y|^{2k - 2\alpha - 1}\, dx\, dy, \quad \text{if } \alpha \notin \mathbb{N}_0 \qquad (2.1) \\ &\sim N^{-2\alpha+1} 2\pi f_*(0) \int_0^1 h_i^{(k)}(x) h_j^{(k)}(x)\, dx, \quad \text{if } a \in \mathbb{N}_0, \end{aligned}$$

where $k = k_\alpha$.

PROOF. A few of the following steps need very detailed calculations, which we omit. Let $d_{k,t}^{(i)} := \Delta_{k,t}(h_i(t/(N+1)))$. Summation by parts gives

$$\begin{aligned} H_{i,n}(\lambda) &:= \sum_{t=1}^{N} h_i\left(\frac{t}{N+1}\right)\exp(-i\lambda t) \\ &= [\exp(-i\lambda) - 1]^{-k} \sum_{t=-k+1}^{N} d_{k,t}^{(i)} \exp(-i\lambda t). \end{aligned}$$

We obtain for any $\delta > 0$,

$$\begin{aligned} &\sum_{s,t=1}^{N} h_1\left(\frac{s}{N+1}\right) h_2\left(\frac{t}{N+1}\right) c(s-t) \\ &= \int_{-\pi}^{\pi} f(\lambda) H_{1,N}(\lambda) H_{2,N}(-\lambda)\, d\lambda \qquad (2.2) \\ &\sim 2f_*(0)\int_0^{\delta} |\lambda|^{2\alpha} H_{1,N}(\lambda) H_{2,N}(-\lambda)\, d\lambda. \end{aligned}$$

If $k = \alpha \in \mathbb{N}$, this is proportional to

$$\begin{aligned} &f_*(0)\int_{-\pi}^{\pi} \left|2\sin\frac{\lambda}{2}\right|^{2k} H_{1,N}(\lambda) H_{2,N}(-\lambda)\, d\lambda \\ &= 2\pi f_*(0) \sum_{t=-k+1}^{N} d_{k,t}^{(1)} d_{k,t}^{(2)} \\ &\sim 2\pi f_*(0) N^{-2k+1} \int_0^1 h_1^{(k)}(x) h_2^{(k)}(x)\, dx. \end{aligned}$$

If $0 < k - \alpha < 1/2$, (2.2) is proportional to

$$\begin{aligned} &2f_*(0)\sum_{s\neq t} d_{k,s}^{(1)} d_{k,t}^{(2)} N^{2k-2\alpha-1} \int_0^{\infty} |x|^{2\alpha-2k} \cos\left(\frac{s-t}{N+1}x\right) dx \\ &= 2f_*(0)\Gamma(2\alpha - 2k + 1)\cos\left\{(2\alpha - 2k + 1)\frac{\pi}{2}\right\} N^{2k-2\alpha-1} \\ &\quad \times \sum_{s\neq t} d_{k,s}^{(1)} d_{k,t}^{(2)} \left|\frac{s-t}{N+1}\right|^{2k-2\alpha-1} \\ &\sim 2f_*(0)\Gamma(2\alpha - 2k + 1)\cos\left\{(2\alpha - 2k + 1)\frac{\pi}{2}\right\} N^{-2\alpha+1} \\ &\quad \times \int_0^1\int_0^1 h_1^{(k)}(x) h_2^{(k)}(y) |x-y|^{2k-2\alpha-1}\, dx\, dy. \end{aligned}$$

If $1/2 \le k - \alpha < 1$, (2.2) is proportional to

$$4f_*(0)\sum_{s,t} d_{k-1,s}^{(1)} d_{k-1,t}^{(2)} \int_0^{\delta} |\lambda|^{2\alpha-2k+1} \cos\{(s-t)\lambda\} \sin\lambda/2\, d\lambda.$$

Since $2\cos\{(s-t)\lambda\}\sin\lambda/2 = \sin\{(s-t+1/2)\lambda\} - \sin\{(s-t-1/2)\lambda\}$, summation by parts yields

$$(2.3) \qquad = -2f_*(0)\sum_{s,t} d_{k,s}^{(1)} d_{k-1,t}^{(2)} \int_0^{\delta} |\lambda|^{2\alpha-2k+1} \sin\{(s-t+1/2)\lambda\}\, d\lambda.$$

In the case $k-\alpha > 1/2$ this is proportional to

$$-2f_*(0)\sum_{s,t} d_{k,s}^{(1)} d_{k-1,t}^{(2)} N^{2k-2\alpha-2} \int_0^{\infty} |x|^{2\alpha-2k+1} \sin\left(\frac{s-1+1/2}{N+1}x\right) dx$$

$$= -2f_*(0)N^{2k-2\alpha-2}\Gamma(2\alpha-2k+2)\sin\left\{(2\alpha-2k+2)\frac{\pi}{2}\right\}$$

$$\times \sum_{s,t} \operatorname{sgn}\left(s-t+\frac{1}{2}\right) d_{k,s}^{(1)} d_{k-1,t}^{(2)} \left|\frac{s-t+1/2}{N+1}\right|^{2k-2\alpha-2}.$$

In the case $k-\alpha = 1/2$, we get the same result by a slightly different calculation. The last expression is proportional to

$$-2f_*(0)\Gamma(2\alpha-2k+2)\sin\left\{(2\alpha-2k+2)\frac{\pi}{2}\right\}N^{-2\alpha+1}$$

$$\times \int_0^1\int_0^1 \operatorname{sgn}(x-y) h_1^{(k)}(x) h_2^{(k-1)}(x) |x-y|^{2k-2\alpha-2}\, dx\, dy$$

$$= 2f_*(0)\Gamma(2\alpha-2k+1)\cos\left\{(2\alpha-2k+1)\frac{\pi}{2}\right\}N^{-2\alpha+1}$$

$$\times \int_0^1\int_0^1 h_1^{(k)}(y) h_2^{(k-1)}(x) |x-y|^{2k-2\alpha-1}\, dx\, dy. \qquad \square$$

REMARK 2.2. If the φ_j are $(k+1)$-times differentiable with bounded derivative and $w(x) = x^{\beta}(1-x)^{\beta}$ with $\beta \ge \alpha$, the above conditions are fulfilled and, in particular, we get $N^{-\alpha-1/2}$ as the rate of convergence for $\hat{\beta}_N$. If $\beta < \alpha$, this will in general no longer be true. An example is the estimation of the mean by the ordinary average of the data ($\beta = 0$) which is known to not have the optimal rate of convergence for $\alpha > 1/2$. We conjecture that the rate is still $N^{-\alpha-1/2}$ for $\beta > \alpha - 1/2$ and $N^{-\beta-1}$ for $\beta < \alpha - 1/2$. By choosing a large β, one may therefore guard against a possible loss in the rate of convergence. If, for example, $\varphi(x) = x^{\gamma}$, the above remarks apply with $\beta + \gamma$ instead of β.

By choosing the particular weight function $w_{\alpha}(x) := x^{\alpha}(1-x)^{\alpha}$ we now prove that we achieve efficiency in polynomial regression. We put $\hat{\beta}_{N,\alpha} = \hat{\beta}_N^{(w_\alpha)}$ for short.

THEOREM 2.3. *Suppose* $\varphi_j(x) = x^{j-1}$, $j = 1, \dots, p$, *and* $w = w_{\alpha}$. *Then we have*:

(i) $$S_N^{-1} R_N S_N^{-1} \sim \Gamma_N^{-1}$$

as N tends to infinity;

(ii) $$\sqrt{N^{1+2\alpha}}\left(\hat{\beta}_{N,\alpha}-\beta_0\right)\to_{\mathscr{D}}\mathscr{N}(0,\gamma^{-1})$$

with

$$\Gamma_{ij}=\frac{\Gamma(\alpha+i)\Gamma(\alpha+j)}{\Gamma(2\alpha+i)\Gamma(2\alpha+j)}\frac{1}{2\alpha+i+j-1}$$

and $\hat{\beta}_{N,\alpha}$ is efficient.

PROOF. (i) The proof of (i) is extremely technical. We have put several technical lemmata into the Appendix. Let Δ, $M(\alpha)$, $T(\alpha)$ and $P(f(\cdot))$ be $p\times p$ matrices defined by

$$\Delta=\left\{(-1)^{i+j}\binom{i-1}{j-1}\right\}_{i,j=1,\dots,p},$$

$$M(\alpha)=\left\{\frac{1}{2\alpha+i+j-1}\right\}_{i,j=1,\dots,p},$$

$$T(\alpha)=\left\{\frac{\prod_{\ell=1}^{j-1}(\alpha+i-1+\ell)}{\prod_{\ell=1}^{j}(2\alpha+i-1+\ell)}\right\}_{i,j=1,\dots,p},$$

$$P(f(\cdot))=\operatorname{diag}\left\{1,f(1),\dots,\prod_{j=1}^{p-1}f(j)\right\}.$$

In particular, for a constant c,

$$P(c)=\operatorname{diag}\{c^0,\dots,c^{p-1}\}.$$

From Theorem A.3, Lemma A.1(vii), Lemma A.4 and Theorem 2.3 of Yajima (1988), we get with the beta function B:

$$R_N\sim N^{-2\alpha+1}2\pi f_*(0)B(\alpha+1,\alpha+1)\Gamma(2\alpha+1)(2\alpha+1)$$
$$\times P(\alpha+\cdot)P\left(\frac{1}{\cdot}\right)P(-1)\,\Delta M(\alpha)\,\Delta' P(-1)P\left(\frac{1}{\cdot}\right)P(\alpha+\cdot),$$

$$S_N\sim NB(\alpha+1,\alpha+1)(2\alpha+1)\left(\sum_{\ell=1}^{i+j-2}\frac{\alpha+\ell}{2\alpha+1+\ell}\right)_{ij}$$

and

$$\Gamma_n\sim N^{2\alpha+1}\{2\pi f_*(0)\}^{-1}\Gamma B(\alpha+1,\alpha+1)\frac{2\alpha+1}{\Gamma(2\alpha+1)}P(\alpha+\cdot)$$
$$\times P\left(\frac{1}{2\alpha+\cdot}\right)M(\alpha)\left(\frac{1}{2\alpha+\cdot}\right)P(\alpha+\cdot),$$

which means that we have to verify

$$Q = P(\alpha + \cdot)P\left(\frac{1}{\cdot}\right)P(-1)\,\Delta M(\alpha)\,\Delta' P(-1)P\left(\frac{1}{\cdot}\right)P(\alpha + \cdot)$$
$$= T(\alpha)' M(\alpha)^{-1} T(\alpha).$$

Since $\Delta T(\alpha)' = P(\alpha + \cdot)P(1/\cdot)\,\Delta M(\alpha)$ and $\Delta^{-1} = P(-1)\,\Delta P(-1)$ [Lemma A.1(iv)], we obtain

$$T(\alpha)' M(\alpha)^{-1} T(\alpha) = P(-1)\,\Delta Q\,\Delta' P(-1).$$

Since

$$(P(-1)\,\Delta Q)_{ij} = \frac{1}{2\alpha + 1} \sum_{k=0}^{i-1} (-1)^k \binom{i-1}{k} \frac{\binom{\alpha + k}{k}\binom{\alpha + j - 1}{j - 1}}{\binom{2\alpha + j + k}{j - 1 + k}}$$

is symmetric in i and j by Lemma A.5, we obtain

$$T(\alpha)' M(\alpha)^{-1} T(\alpha) = P(-1)\,\Delta Q\,\Delta' P(-1) = Q\,\Delta' P(-1)\,\Delta' P(-1) = Q.$$

(ii) $\hat{\beta}_N^{(w_\alpha)}$ is already Gaussian with mean β and asymptotic variance Γ^{-1}. □

REMARK 2.4. From the above result one can easily deduce efficiency for polynomial regression when the regressors are of the form $X_i = (i^0, \dots, i^{P-1})$. This was the situation considered by Yajima (1988). In this case one has to multiply the matrices Σ_N, S_N and Γ_N from the left and the right by the matrix $P(N) = \operatorname{diag}\{N^0, \dots, N^{p-1}\}$, which leads to the central limit theorem

$$\sqrt{N^{1+2\alpha}}\, P(N)\left(\hat{\beta}_{N,\alpha} - \beta_0\right) \to_{\mathscr{D}} \mathscr{N}(0, \Gamma^{-1}).$$

We now extend the above result to more general regressors. Let $g_1(x)$, $g_2(x), \dots$ be the orthonormal polynomials with respect to the inner product $(g, h) = \int_0^1 g(x)h(x)w_\alpha(x)\,dx$ obtained from $1, x, x^2, \dots$ by the Gram–Schmidt procedure. We mention that

$$g_{m+1}(x) = \text{const}_{\alpha, m} \sum_{\ell=0}^{m} \binom{m + \alpha}{\ell}\binom{m + \alpha}{m - \ell}(-1)^{m-\ell}(1 - x)^{m-\ell} x^{\ell},$$
$$m = 0, \dots,$$

which are (up to a substitution) the Jacobi polynomials.

THEOREM 2.5. *Suppose the* $\varphi_j(x)$, $j = 1, \dots, p$, *span the same linear space as p Jacobi polynomials* $g_{i_1}, \dots, g_{i_p}$. *Let* $w = w_\alpha$.

(i) *Then we obtain, for the corresponding matrices,*

$$S_N^{-1} \Sigma_N S_N^{-1} \sim \Gamma_N^{-1}.$$

(ii) $\hat{\beta}_{N,\alpha}$ *then is asymptotically normal with rate* $N^{-\alpha-1/2}$ *and efficient.*

PROOF. We start with a preliminary consideration. Let $\ell \in \mathbb{N}$ and A_ℓ be an $\ell \times \ell$ matrix with

$$(g_1, \dots, g_\ell)' = A'_\ell(1, \dots, x^{\ell-1})'.$$

Furthermore, let U_ℓ and V_ℓ be the design matrices corresponding to the regressors $(g_1, \dots, g_\ell)$ and $(1, \dots, x^{\ell-1})$, respectively. Then we have $U_\ell = V_\ell A_\ell$ and from Theorem 2.3,

$$(V'_\ell I_w V_\ell)^{-1}(V'_\ell I_w \Sigma I_w V_\ell)(V'_\ell I_w V_\ell)^{-1} \sim (V'_\ell \Sigma^{-1} V_\ell)^{-1}.$$

which implies

$$(U'_\ell I_w U_\ell)^{-1}(U'_\ell I_w \Sigma I_w U_\ell)(U'_\ell I_w U_\ell)^{-1} \sim (U'_\ell \Sigma^{-1} U_\ell)^{-1}.$$

Since $U'_\ell I_w U_\ell$ is approximately the identity matrix, we get

$$U'_\ell I_w \Sigma I_w U_\ell \sim (U'_\ell \Sigma^{-1} U_\ell)^{-1}.$$

Since this holds for all ℓ we can conclude from Theorem 8.2.1.1(f) of Graybill (1983) that $U'_\ell I_w \Sigma I_w U_\ell \sim D$ with some diagonal matrix D.

Suppose now that X and Z are the design matrices corresponding to the regressors φ_j and g_{ij}, $j = 1, \dots, p$, respectively, that is, $X = ZB$ with some regular matrix $p \times p$ matrix B. From the above consideration we obtain (for the diagonal matrices)

$$Z' I_w \Sigma I_w Z \sim (Z' \Sigma^{-1} Z)^{-1}$$

and with $Z' I_w Z \sim I$,

$$(X' I_w X)^{-1}(X' I_w \Sigma I_w X)(X' I_w X)^{-1} \sim (X' \Sigma^{-1} X)^{-1},$$

which implies (i). As in the proof of Theorem 2.3, part (ii) follows immediately. □

It seems to be more natural to prove Theorem 2.5 first and then get the polynomial regression as a special case. However, the effort would be even larger, since in our proof we could make use of Theorem 2.3 of Yajima (1988) for the polynomial case.

REMARK 2.6. (i) We conjecture that the above weighted least squares estimator $\hat{\beta}_{N,\alpha}$ is not efficient for trigonometric design. It seems that in this case arguments similar to the ones given in the Introduction can be made rigorous, which lead to a relative efficiency of

$$\int_0^1 w(x)^2\,dx \bigg/ \left(\int_0^1 w(x)\,dx\right)^2 > 1 \quad \text{if } w \neq 1.$$

(ii) One might conjecture that efficiency for $\hat{\beta}_{N,\alpha}$ holds whenever $F_x(\lambda) = c\chi_{[0,\pi]}(\lambda)$ (cf. Section 1). We give a counterexample for this: Suppose that we have only one regressor $\varphi_1(x) = x^m$ with $m \neq 0$. Then the linear space spanned by φ_1 does not coincide with the space spanned by any of the g_j. In fact, the corresponding weighted least squares estimator $\hat{\beta}_{N,\alpha}$ is no longer efficient. This follows since the corresponding values of R_N, S_N and Γ_N are the $(m+1, m+1)$ entries of the matrices listed in the proof of Theorem 2.3. For $m = 1$ we get, for example,

$$S_N^{-1} R_N S_N^{-1} \Gamma_N \sim \frac{4(\alpha+1)^3}{(\alpha+2)^2(2\alpha+1)},$$

which is not equal to 1 in general. This is a bit surprising because we know from Theorem 2.3 that the regression with $(1, x)$ is efficient.

(iii) We now try to make the above results more transparent with the following heuristics: Suppose $w = w_\alpha$ and X is a $n \times p$ design matrix with $\Sigma_N^{-1} X \sim I_w X$. One easily checks that this implies

$$S_N^{-1} R_N S_N^{-1} \sim \Gamma_N^{-1}, \tag{2.4}$$

that is, we have efficiency of $\hat{\beta}_{N,\alpha}$ for this design. More generally, we can derive (2.4) if $\Sigma_N^{-1} X \sim I_w XA$ with a $p \times p$ matrix A of full rank. This means that we obtain efficiency if $\Sigma_N^{-1} X$ and $I_w X$ span (approximately) the same space. The above example now can be explained as follows: $\Sigma_N^{-1}(1, X_i)_{i=1,\dots,N}$ and $I_w(1, X_i)_{i,\dots,N}$ $(X_i = i/N)$ span approximately the same space, while $\Sigma_N^{-1}(X_i)_i$ and $I_w(X_i)_i$ do not. Although we could check this for a specific case $(\alpha = 1)$, it is in general not clear how to make this heuristic rigorous (in particular, the correct definition of the " $\sim$ " sign is not clear).

We now construct fully efficient weighted M-estimates. It was observed by Taqqu (1975) that the asymptotic behaviour of the partial sums of long range dependent random variables depends only on the behaviour of the first nonzero term in a Hermite expansion. Beran and Künsch (1985) and Beran (1991) used this result to prove that the ordinary mean and M-estimators for location have the same relative efficiency. It is not difficult to generalize these results to our situation.

Let

$$H_n(x) = (-1)^n \exp\left(\frac{x^2}{2}\right) \frac{d^n}{(dx)^n} \exp\left(\frac{-x^2}{2}\right), \qquad n \in \mathbb{N}_0$$

be the Hermite polynomials. They form a complete orthogonal system in $\mathcal{L}_2(\mathbb{R}, \varphi)$ with $\varphi(x) = (1/\sqrt{2\pi})\exp(-x^2/2)$ with $\mathbb{E} H_n(X) H_v(X) = \delta_{nv} n!$, where $X \sim \mathcal{N}(0,1)$.

The series $\sum_{n=0}^\infty (J(n)/n!) H_n(x)$ with $J(n) = \int G(x) H_n(x) \varphi(x)\, dx$ converges in $\mathcal{L}_2(\mathbb{R}, \varphi)$ to $G(x)$. In particular, $H_0(x) = 1$ and $H_1(x) = x$. Furthermore,

$$\mathbb{E} H_n\left(\frac{\varepsilon_i}{\sigma}\right) H_v\left(\frac{\varepsilon_j}{\sigma}\right) = \delta_{nv} n! \frac{c(i-j)^n}{\sigma^{2n}},$$

where $\sigma^2 = \mathbb{E}\varepsilon_i^2$ [cf. Feller (1971)].

Consider now the weighted M-estimate $\tilde{\beta}_{N,\alpha}$, defined as the solution of the equation

$$\sum_{i=1}^N w_\alpha\left(\frac{i}{N+1}\right) X_i' \psi(Y_i - X_i \beta) = 0. \tag{2.5}$$

THEOREM 2.7. *Suppose ψ is a nondecreasing, differentiable function with $\mathbb{E}\psi(\varepsilon_i) = 0$, $\mathbb{E}\psi'(\varepsilon_i) \neq 0$ and $|\psi'(\varepsilon_i + \delta_n) - \psi'(\varepsilon_i)| \to_p 0$ for $\delta_n \to_p 0$. Let further $\varphi_1, \dots, \varphi_p$ be a set of continuous functions such that the matrix $\int_0^1 \varphi_i(x)\varphi_j(x)w_\alpha(x)\,dx$ is regular. Then*:

(i) *There exists a sequence $\tilde{\beta}_{N,\alpha}$ that solves (2.5) with $\tilde{\beta}_{N,\alpha} \to \beta_0$ a.s.*

(ii)
$$\sqrt{N^{1+2\alpha}}\left(\tilde{\beta}_{N,\alpha} - \hat{\beta}_{N,\alpha}\right) \to_p 0$$

and $\tilde{\beta}_{N,\alpha}$ is asymptotically normal.

(iii) *If in addition the $\varphi_1, \dots, \rho_p$ span the same linear space as p Jacobi polynomials $g_{i_1}, \dots, g_{i_p}$, then $\tilde{\beta}_{N,\alpha}$ is also efficient.*

PROOF. The proof is given in Dahlhaus (1992).

REMARK 2.8. The condition $\mathbb{E}\psi'(\varepsilon_i) \neq 0$ implies that ψ is of Hermite rank 1. We do not consider M-functions of arbitrary rank since the typical M-function for the location estimation is symmetric around zero, which implies $\mathbb{E}\psi'(\varepsilon_i) \neq 0$. Furthermore, this condition is also necessary to obtain the efficiency result.

3. Adaptive weighted least squares estimators.

In this section we prove that the weighted least squares estimator remains efficient when the weighting function w_α is replaced by $w_{\hat{\alpha}_N}$, where $\hat{\alpha}_N$ is an N^ρ-consistent estimator of α with some (arbitrarily small) $\rho > 0$.

THEOREM 3.1. *Let $\varphi_1, \dots, \varphi_p$ be a set of continuous functions such that the matrix $\{\int_0^1 \varphi_i(x)\varphi_j(x)w_\alpha(x)\,dx\}_{ij}$ is regular and let $\hat{\alpha} = \hat{\alpha}_N$ be an N^ρ-consistent estimator of α for some $\rho > 0$. Then $\sqrt{N^{1+2\alpha}}(\hat{\beta}_{N,\hat{\alpha}} - \hat{\beta}_{N,\alpha}) \to_p 0$ and $\hat{\beta}_{N,\hat{\alpha}}$ is asymptotically normal. If the φ_j, $j = 1, \dots, p$, span the same space as p Jacobi polynomials $g_{i_1}, \dots, g_{i_p}$, then $\hat{\beta}_{N,\hat{\alpha}}$ is also efficient.*

PROOF. We use a chaining argument. Let $\alpha > -1/2$ and $\varepsilon > 0$ with $\alpha - \varepsilon > 1/2$. Let $\delta > 0$ and $\{\delta_j\}$ be a positive sequence with $\delta = \sum_{j=0}^\infty \delta_j$. For each $\gamma \in [\alpha, \alpha + \varepsilon]$ there exists a sequence $\gamma_j \to \gamma$ of the form $\gamma_j = \alpha + \varepsilon(n_j/2^j)$, $n_j \le 2^j$, with $\gamma_j = \gamma_{j+1}$ or $|\gamma_j - \gamma_{j+1}| \le \varepsilon/2^{j+1}$ (dyadic representation). This implies, with $\gamma_{ij} = \alpha + \varepsilon(i/2^j)$, $\delta_{ij} = \gamma_{ij} + \varepsilon/2^{j+1}$,

$$\mathbb{P}\left(\sup_{\gamma \in [\alpha, \alpha+\varepsilon]} \sqrt{N^{1+2\alpha}}\,|\hat{\beta}_{N,\gamma} - \hat{\beta}_{N,\alpha}| \ge \delta\right)$$

$$\le \sum_{j=0}^{\infty} \sum_{i=0}^{2^j - 1} \mathbb{P}\left(\sqrt{N^{1+2\alpha}}\,|\hat{\beta}_{N,\gamma_{ij}} - \hat{\beta}_{N,\delta_{ij}}| \ge \delta_j\right) \tag{3.1}$$

$$\le N^{1+2\alpha} \sum_{j=0}^{\infty} 1/\delta_j^2 \sum_{i=0}^{2^j-1} \mathbb{E}\left(\hat{\beta}_{N,\gamma_{ij}} - \hat{\beta}_{N,\delta_{ij}}\right)'\left(\hat{\beta}_{N,\gamma_{ij}} - \hat{\beta}_{N,\delta_{ij}}\right).$$

The expectation is equal to (put $\gamma = \gamma_{ij}$, $\delta = \delta_{ij}$, $I_\gamma = I_{w_\gamma}$)

$$\operatorname{tr}\left\{\left[(X'I_\gamma X)^{-1}X'I_\gamma - (X'I_\delta X)^{-1}X'I_\delta\right]\Sigma\left[I_\gamma X(X'I_\gamma X)^{-1} - I_\delta X(X'I_\delta X)^{-1}\right]\right\},$$

which can be estimated with a mean value $\eta \in [\gamma, \delta]$ and $J_\eta := I_{(\partial/\partial\alpha)w_\alpha|_{\alpha=\eta}}$ by

$$\begin{aligned}(\gamma-\delta)^2 2\operatorname{tr}&\left\{(X'I_\eta X)^{-1}(X'J_\eta\Sigma J_\eta X)(X'I_\eta X)^{-1}\right\}(\gamma-\delta)^2\\ &\times 2\operatorname{tr}\left\{(X'I_\eta X)^{-1}(X'J_\eta X)(X'I_\eta X)^{-1}\right.\\ &\qquad\left.\times(X'I_\eta\Sigma I_\eta X)(X'I_\eta X)^{-1}(X'J_\eta X)(X'I_\eta X)^{-1}\right\},\end{aligned}$$

which is less than

$$\begin{aligned}&2(\gamma-\delta)^2 p\|(X'I_\eta X)^{-1}\|^2\|X'J_\eta\Sigma J_\eta X\|\\ &\quad+2(\gamma-\delta)^2 p\|(X'I_\eta X)^{-1}\|^4\|X'J_\eta X\|^2\|X'I_\eta\Sigma I_\eta X\|,\end{aligned}$$

where $\|A\|$ is the spectral norm of A which, in the case of a symmetric matrix, is equal to the largest absolute root of A or equal to $1/C$ (smallest absolute root of A^{-1}). Straightforward calculations show that

$$\|(X'I_\eta X)^{-1}\| \le KN^{-1} \quad \text{and} \quad \|X'J_\eta X\| \le KN$$

uniformly for $\eta \in [\alpha, \alpha+\varepsilon]$. Furthermore,

$$\begin{aligned}\|X'I_\eta\Sigma I_\eta X\| &= \sup_{\|y\|=1}\sum_{l,m=1}^{P} y_l y_m \sum_{s,t=1}^{N}\varphi_l\left(\frac{s}{N+1}\right)\varphi_m\left(\frac{t}{N+1}\right)\\ &\qquad\times w_\eta\left(\frac{s}{N+1}\right)w_\eta\left(\frac{t}{N+1}\right)c(s-t)\\ &= \sup_{\|y\|=1}\int f(\gamma)\left|\sum_{s=1}^{N}\left[\sum_l y_l\varphi_l\left(\frac{s}{N+1}\right)\right]\right.\\ &\qquad\qquad\left.\times w_\eta\left(\frac{s}{N+1}\right)\exp(i\gamma s)\right|^2 d\gamma.\end{aligned}$$

Since $h_y(x) := \sum_l y_l\varphi_l(x)w_\eta(x)$ is of smoothness α, it follows from Theorem 2.1 that

$$\|X'I_\eta\Sigma I_\eta X\| \le KN^{-2\alpha+1}$$

and similarly

$$\|X'J_\eta\Sigma J_\eta X\| \le K_v N^{-2\alpha+1+v}$$

with some arbitrarily small $v > 0$. Choose $v = \rho$. This gives as an upper bound of (3.1),

$$KN^{\rho} \sum_{j=0}^{\infty} \frac{1}{\delta_j^2} \sum_{i=1}^{2^j-1} (\gamma_{ij} - \delta_{ij})^2$$

$$\le KN^{\rho}\varepsilon^2 \sum_{j=0}^{\infty} \left(\delta_j^2 2^{j+1}\right)^{-1} \le KN^{\rho}\varepsilon^2,$$

if one chooses, for example, $\delta_j \sim \text{constant}/j^2$. The analogous expression for $\sup_{\gamma \in [\alpha-\varepsilon, \alpha]}$ is estimated similarly with one exception, namely, for $\eta < \alpha$ with h_y as above,

$$\|X' I_\eta \Sigma I_\eta X\| = \sup_{\|y\|=1} \int f(\gamma) \left| \sum_{s=1}^{N} h_y\left(\frac{s}{N+1}\right) \exp(i\gamma s) \right|^2 d\gamma,$$

since $f(\gamma) \le K|\gamma|^{2\alpha} \le K|\gamma|^{2\eta}$ this is, with Theorem 2.1, bounded by

$$KN^{-2\eta+1} \le KN^{-2\alpha+1+2\varepsilon} \left(K_\rho N^{-2\alpha+1+2\varepsilon+\rho} \text{ for } J_\eta \text{ instead of } I_\eta\right).$$

Choose now $\varepsilon = \varepsilon'/N^{\rho}$, which gives the result. □

REMARK 3.2. One easily checks that the above proof holds for polynomial regression with $X_i = (i^0, \dots, i^{p-1})$. One then obtains

$$\sqrt{N^{1+2\alpha}} P(N)\left(\hat{\beta}_{N,\hat{\alpha}} - \hat{\beta}_{N,\alpha}\right) \to_P 0.$$

The above result can be used in an iteration. One starts with the least squares estimator for β (with the ordinary mean in the location problem) which is $\sqrt{N^{1+2\alpha}}$-consistent, estimates α from the residuals which is $\sqrt{N}$ consistent [cf. Yajima (1988), Theorem 3.2] and then reestimates β with a weighted regression (or with a weighted mean) with weight function $w_{\hat{\alpha}}$. If α is possibly larger than 1/2 one may start with a weighted regression with a sufficiently smooth weight function (cf. Remark 2.2).

APPENDIX

This Appendix provides the technical details for the proof of Theorem 2.3.

LEMMA A.1.

(i) $\Delta_{k,i}(f(i)) = \sum_{j=0}^{k} \binom{k}{j} (-1)^j f(i-j),$

(ii) $\Delta_{k,n} f(\ell + n)|_{n=0} = (-1)^k \Delta_{k,n} f(-k + \ell - n)|_{n=0},$

(iii) $\Delta(a_1, \dots, a_p)' = (\Delta_{i-1,i}(a_i))_{i=1,\dots,p}, \qquad a_i \in \mathbb{R},$

(iv) $\Delta^{-1} = \left\{ \binom{i-1}{j-1} \right\}_{i,j=1,\dots,N} = P(-1)\,\Delta P(-1),$

(v) $\Delta_{k,n}\left(\frac{1}{x+n}\right)\Big|_{n=0} = \frac{(-1)^k k!}{\prod_{\ell=0}^{k}(x-\ell)}$ *for all* x *with* $x,\dots,x-k \neq 0$,

(vi) $$\Delta_{k,\ell}\left(\prod_{j=1}^{\ell} \frac{\alpha+n+j}{2\alpha+v+1+j}\right)$$
$$= (-1)^k \left(\prod_{j=1}^{\ell-k} \alpha+n+j\right)\left(\prod_{j=1}^{k} \alpha+v-n+j\right)$$
$$\times \left(\prod_{j=1}^{\ell} \frac{1}{2\alpha+v+1+j}\right)$$ *for all* $k \leq \ell$ *and* $2\alpha+v+2>0$,

(vii) $\Delta M(\alpha)\Delta' = \frac{(-1)^{i+j-2}(i+j-2)!}{\prod_{\ell=0}^{i+j-2}(2\alpha+1+\ell)},$

(viii) $\Delta T(\alpha)' = P(\alpha+\cdot)P\left(\frac{1}{\cdot}\right)\Delta M.$

PROOF. The proofs of (i)–(iii) are straightforward. To prove (iv), note that

$$\sum_{k=j}^{i} (-1)^{i+k}\binom{i-1}{k-1}\binom{k-1}{j-1} = \delta_{ij}.$$

Identities (v) and (vi) are proved by induction over k; (vii) follows from (iii) and (v) since

$$\begin{aligned}(\Delta M(\alpha)\Delta')_{ij} &= (\Delta(\Delta M(\alpha))')_{ji} = \Delta_{j-1,j}(\Delta M(\alpha))_{ij}\\ &= \Delta_{j-1,j}\Delta_{i-1,i}\frac{1}{2\alpha+i+j-1}\\ &= \Delta_{i+j-2,n}\frac{1}{2\alpha+i+j+n-1}\Big|_{n=0}\\ &= \frac{(-1)^{i+j-2}(i+j-2)!}{\prod_{\ell=0}^{i+j-2}(2\alpha+i+j-1-\ell)}.\end{aligned}$$

To prove (viii) we note that

$$(\Delta T(\alpha)')_{ij} = \frac{1}{2\alpha+j}\Delta_{i-1,i-1}\left(\prod_{\ell=1}^{i-1}\frac{\alpha+j-1+\ell}{2\alpha+j+\ell}\right).$$

Therefore, by using (vi), (v) and (iii) we get

$$\begin{aligned}\Delta T(\alpha)' &= P(\alpha + \cdot)P\left(\frac{1}{\cdot}\right)\left\{\frac{(-1)^{i-1}(i-1)!}{\prod_{\ell=0}^{i-1} 2\alpha + j + \ell}\right\}_{ij} \\ &= P(\alpha + \cdot)P\left(\frac{1}{\cdot}\right)\Delta M(\alpha). \qquad \square\end{aligned}$$

LEMMA A.2.

(i) $\displaystyle\sum_{\ell=0}^{\infty}\left(\frac{1}{\ell-\alpha} - \frac{1}{\ell+1+\alpha}\right) = -\pi\cot(\alpha\pi)$ *for all* $\alpha \notin \mathbb{Z}$,

(ii) $$\sum_{\ell=0}^{\infty}\frac{1}{\ell+\alpha+1-u}\,\frac{1}{\ell-\alpha+v} = -\frac{1}{2\alpha-u-v+1}\left[\pi\cot(\alpha\pi) + \begin{cases}\displaystyle\sum_{v}^{u-1}\frac{1}{\alpha-\ell}, & u > v\\ 0, & u = v\\ \displaystyle\sum_{u}^{v-1}\frac{1}{\alpha-\ell}, & u < v\end{cases}\right]$$

for all α, u and v with $2\alpha - u - v + 1 \neq 0$ and $\alpha + u$, $\alpha + v \notin \mathbb{Z}$.

PROOF. (i) For $-1 < \alpha < 0$ we obtain

$$\sum_{\ell=0}^{\infty}\left(\frac{1}{\ell-\alpha} - \frac{1}{\ell+1+\alpha}\right) = \int_0^1 \frac{x^{-\alpha-1} - x^{\alpha}}{1-x}\,dx = -\pi\cot(\alpha\pi).$$

For all other α the result is obtained by shifting the sum.

(ii) We obtain, for example, for $u > v$,

$$\begin{aligned}&\sum_{\ell=0}^{\infty}\frac{1}{\ell+\alpha+1-u}\,\frac{1}{\ell-\alpha+v}\\ &\quad= \frac{1}{2\alpha-u-v+1}\sum_{\ell=0}^{\infty}\left(\frac{1}{\ell-\alpha+v} - \frac{1}{\ell+\alpha+1-u}\right)\\ &\quad= \frac{1}{2\alpha-u-v+1}\left[\sum_{\ell=0}^{\infty}\left(\frac{1}{\ell-\alpha+v} - \frac{1}{\ell+\alpha+1-v}\right) - \sum_{\ell=v}^{u-1}\frac{1}{\alpha-\ell}\right].\end{aligned}$$

The result now follows from (i). $\square$

We now calculate the coefficients of $\Sigma_N = X'I_w\Sigma I_w X$ for polynomial regression $[\varphi_j(x) = x^{j-1}]$ and the particular weight function $w_\alpha(x) = x^\alpha(1-x)^\alpha$. This means that we have to calculate the expressions from Theorem 2.1 for $h_j(x) = \varphi_j(x)w_\alpha(x) = x^{\alpha+j-1}(1-x)^\alpha$. All h_j are of smoothness α.

THEOREM A.3. *Let* $\alpha \in (-1/2, \infty)$, $\varphi_j(x) = x^{j-1}$ *and* $w(x) = w_\alpha(x) = x^\alpha(1 - x)^\alpha$. *Then*

$$R_{N_{ij}} \sim 2\pi f_*(0) N^{-2\alpha+1} \frac{\Gamma(\alpha+1)^2}{2\alpha+1} \frac{\binom{\alpha+i-1}{i-1}\binom{\alpha+j-1}{j-1}}{\binom{2\alpha+i+j-1}{i+j-2}}.$$

PROOF. We calculate the expressions of Theorem 2.1. We consider only the more difficult case $\alpha \notin \mathbb{N}_0$. The case $\alpha \in \mathbb{N}_0$ is treated similarly. We split the integral in (2.1) into the domains $\{x \geq y\}$ and $\{x < y\}$. Since

$$h_i^{(k)}(x) = k! \sum_{m=0}^{k} (-1)^{k-m} \binom{\alpha}{k-m} \binom{\alpha+i-1}{m} x^{\alpha+i-1-m} (1-x)^{\alpha-k-m},$$

we obtain with the substitution $z = (1-x)/(1-y)$:

$$\begin{aligned}
&\int_y^1 h_i^{(k)}(x)(x-y)^{2k-2\alpha-1}\,dx \\
&\quad = k! \sum_{m=0}^{k} (-1)^{k-m} \binom{\alpha}{k-m}\binom{\alpha+i-1}{m} (1-y)^{k-\alpha+m} \\
&\qquad \times \int_0^1 z^{\alpha+m-k}(1-z)^{2k-2\alpha-1}\{1-(1-y)z\}^{\alpha+i-1-m}\,dz \\
&\quad = k! \sum_{m=0}^{k}\sum_{\ell=0}^{\infty} (-1)^{k-m+\ell}(1-y)^{k-\alpha+\ell+m}\binom{\alpha}{k-m}\binom{\alpha+i-1}{m} \\
&\qquad \times \binom{\alpha+i-1-m}{\ell} B(\alpha+\ell+m-k+1, 2k-2\alpha).
\end{aligned}$$

After a change of summation ($\ell + m \mapsto \ell$) this is equal to [note that $\binom{\alpha+i-1-m}{\ell-m}\binom{\alpha+i-1}{m} = \binom{\alpha+i-1}{\ell}\binom{\ell}{m}$]

$$\sum_{\ell=0}^{\infty} (-1)^{k+\ell}(1-y)^{k+\ell-\alpha}\binom{\alpha+i-1}{\ell}\binom{\ell+\alpha}{k} k! B(\alpha+\ell-k+1, 2k-2\alpha).$$

In order to obtain the integral in (2.1) we calculate

$$\begin{aligned}
\int_0^1 h_j^{(k)}(y)(1-y)^{k+\ell-\alpha}\,dy &= k! \sum_{n=0}^{k} (-1)^{k-n}\binom{\alpha}{k-n}\binom{n+\ell}{n} B(\alpha+j, \ell+1) \\
&= (-1)^k \binom{\alpha-\ell-1}{k} k! B(\alpha+j, \ell+1)
\end{aligned}$$

and therefore

$$\int_0^1 \int_y^1 h_j^{(k)}(y) h_i^{(k)}(x) |x-y|^{2k-2\alpha-1}\,dx\,dy$$

$$= \sum_{\ell=0}^{\infty} (-1)^{\ell} \binom{\alpha+i-1}{\ell} \binom{\ell+\alpha}{k} k! \binom{\alpha-\ell-1}{k} k! B(\alpha+j,\ell+1)$$

$$\times B(\alpha+\ell-k+1, 2k-2\alpha)$$

$$= \frac{\Gamma(\alpha+1)\Gamma(2k-2\alpha)}{\Gamma(-\alpha)} \left\{ \prod_{u=1}^{i-1} (\alpha+u) \right\} \left\{ \prod_{v=1}^{j-1} (\alpha+v) \right\} (-1)^{i+k-1}$$

$$\times \sum_{\ell=0}^{\infty} \frac{1}{\{\prod_{u=0}^{j-1}(\ell+\alpha+j-u)\}\{\prod_{v=0}^{i-1}(\ell-\alpha-v)\}}.$$

By Lemma A.1 we get

$$\frac{1}{\prod_{u=0}^{j-1}(\ell+\alpha+j-u)} = \frac{(-1)^{j-1}}{(j-1)!} \Delta_{j-1,u} \left(\frac{1}{\ell+\alpha+j-u} \right) \Bigg|_{u=0}$$

$$= \frac{1}{(j-1)!} \Delta_{j-1,u} \left(\frac{1}{\ell+\alpha+1-u} \right) \Bigg|_{u=0},$$

and therefore the sum in the above expression is equal to

$$\sum_{\ell=0}^{\infty} \frac{1}{(j-1)!} \frac{(-1)^{i-1}}{(i-1)!} \Delta_{j-1,u} \left(\frac{1}{\ell+\alpha+1-u} \right) \Bigg|_{u=0} \Delta_{i-1,v} \left(\frac{1}{\ell-\alpha+v} \right) \Bigg|_{v=0}$$

$$= \frac{1}{(j-1)!} \frac{(-1)^{i-1}}{(i-1)!} \Delta_{j-1,u} \left[\Delta_{i-1,v} \left(\sum_{\ell=0}^{\infty} \frac{1}{\ell+\alpha+1-u} \frac{1}{\ell-\alpha+v} \right) \right] \Bigg|_{u,v=0}.$$

Together with the corresponding integral over $\{x < y\}$ (exchange i and j!) and Lemma A.2(ii) the whole expression in (2.1) now becomes equal to

$$2 f_*(0) N^{-2\alpha+1} \Gamma(2\alpha-2k+1) \cos\left\{ (2\alpha-2k+1)\frac{\pi}{2} \right\} \frac{\Gamma(\alpha+1)\Gamma(2k-2\alpha)}{\Gamma(-\alpha)}$$

$$\times (-1)^{k-1} \binom{\alpha+i-1}{i-1} \binom{\alpha+j-1}{j-1}$$

$$\times 2\pi \cot(\alpha\pi) \Delta_{i-1,u} \left[\Delta_{j-1,v} \left(\frac{1}{2\alpha-u-v+1} \right) \right]$$

[the remainder terms from Lemma A.2(ii) cancel for the two integrals]. By using Lemma A.1(v) and the relations

$$-2\pi \cot(\alpha\pi) = \frac{\Gamma(\alpha)\Gamma(1-\alpha)\Gamma(-\alpha)\Gamma(1+\alpha)}{\Gamma(2\alpha)\Gamma(1-2\alpha)},$$

$$\frac{1}{\pi} \cos\left\{ (2\alpha-2k+1)\frac{\pi}{2} \right\} = (-1)^{k-1} \frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)},$$

we obtain the result for the case $\alpha \notin \mathbb{N}_0$. The proof for $\alpha \in \mathbb{N}_0$, which uses similar considerations, is omitted. □

LEMMA A.4. *Let* $\alpha \in (-1/2, \infty)$, $\varphi_j(x) = x^{j-1}$ *and* $w(x) = w_\alpha(x) = x^\alpha(1-x)^\alpha$. *Then*

$$\begin{aligned} S_{N_{ij}} &\sim NB(i+j-1+\alpha, \alpha+1) \\ &= NB(\alpha+1, \alpha+1) \prod_{\ell=1}^{i+j-2} \frac{\alpha+\ell}{2\alpha+1+\ell}. \end{aligned}$$

The proof is trivial.

LEMMA A.5. *For* $\alpha > -1/2$ *and* $i, j \in \mathbb{N}$, *we have*

$$\begin{aligned} &\sum_{k=0}^{i-1} (-1)^k \binom{i-1}{k} \frac{\binom{\alpha+k}{k}\binom{\alpha+j-1}{j-1}}{\binom{2\alpha+j+k}{j-1+k}} \\ &\quad = \sum_{v=0}^{\min(i-1,j-1)} (-1)^v \binom{i-1}{v}\binom{j-1}{v} \\ &\qquad \times \left\{\prod_{\ell=1}^{v} (\alpha+\ell)\right\}\left\{\prod_{\ell=1}^{i+j-2-v} (\alpha+\ell)\right\}. \end{aligned} \tag{A.1}$$

PROOF. (A.1) is equal to

$$\prod_{\ell=1}^{j-1} (\alpha+\ell) \sum_{k=0}^{i-1} (-1)^k \binom{i-1}{k}\binom{j-1+k}{k} \frac{\prod_{\ell=1}^{k}(\alpha+\ell)}{\prod_{\ell=1}^{j-1+k}(2\alpha+1+\ell)}.$$

Since

$$\binom{j-1+k}{k} = \sum_{v=0}^{\min(j-1,k)} \binom{j-1}{v}\binom{k}{v} \quad \text{and}$$

$$\binom{i-1}{k}\binom{k}{v} = \binom{i-1}{v}\binom{i-1-v}{k-v},$$

this is equal to

$$\begin{aligned} &\frac{\prod_{\ell=1}^{j-1}(\alpha+\ell)}{\prod_{\ell=1}^{j-1}(2\alpha+1+\ell)} \sum_v (-1)^v \binom{i-1}{v}\binom{j-1}{v}(-1)^{i-1-v} \\ &\quad \times \Delta_{i-1-v,i-1}\left(\prod_{\ell=1}^{i-1} \frac{\alpha+\ell}{2\alpha+j+\ell}\right), \end{aligned}$$

from which we obtain the result by using Lemma A.1(vi). □

REFERENCES

ADENSTEDT, R. K. (1974). On large-sample estimation for the mean of a stationary random sequence. *Ann. Statist.* **2** 1095–1107.

BERAN, J. (1991). *M* estimators of location for data with slowly decaying serial correlations. *J. Amer. Statist. Assoc.* **86** 704–708.

BERAN, J. and KÜNSCH, H. (1985). Location estimation for processes with long range dependence. Research Report 40, Seminar für Statistik, ETH, Zürich.

DAHLHAUS, R. (1992). Efficient location and regression estimation for long range dependent regression models. Preprint 660, SFB 123 Stochastische mathematische Modelle, Univ. Heidelberg.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications.* Wiley, New York.

GRAYBILL, F. A. (1983). *Matrices with Applications in Statistics*, 2nd ed. Wadsworth, Belmont, CA.

GRENANDER, U. (1954). On estimation of regression coefficients in the case of an autocorrelated disturbance. *Ann. Math. Statist.* **25** 252–272.

HAMPEL, F. (1987). Data analysis and self-similar processes. *Bull. Inst. Internat. Statist.* **52** 235–254.

KOUL, H. L. (1992). *M*-estimators in linear models with long range dependent errors. *Statist. Probab. Lett.* **14** 153–164.

KOUL, H. L. and MUKHERJEE, K. (1993). Asymptotics of R-, MD- and LAD-estimators in linear regression models with long range dependent errors. *Probab. Theory Related Fields* **95** 535–553.

MILLAR, P. W. (1983). The minimax principle in asymptotic statistical theory. *Ecole d'Eté de Probabilitiés de Saint-Flour XI. Lecture Notes in Math.* **976**. Springer, Berlin.

SAMAROV, A. and TAQQU M. S. (1988). On the efficiency of the sample mean in long-memory noise. *J. Time Ser. Anal.* **9** 192–200.

TAQQU, M. S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrsch. Verw. Gebiete* **31** 287–302.

YAJIMA, Y. (1988). On estimation of a regression model with long-memory stationary errors. *Ann. Statist.* **16** 791–807.

YAJIMA, Y. (1991). Asymptotic properties of the LSE in a regression model with long-memory stationary errors. *Ann. Statist.* **19** 158–177.

UNIVERSITÄT HEIDELBERG
INSTITUT FÜR ANGEWANDTE MATHEMATIK
IM NEUENHEIMER FELD 294
D-69120 HEIDELBERG
GERMANY