Journal of Statistical Planning and Inference 140 (2010) 3742–3753

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Goodness-of-fit testing under long memory

Hira L. Koul [a,1], Donatas Surgailis [b,*,2]

[a] Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824-1027, USA
[b] Vilnius Institute of Mathematics and Informatics, 08663 Vilnius, Lithuania

ARTICLE INFO

Available online 21 May 2010

MSC:
primary, 62M09;
secondary, 62M10, 62M99

Keywords:
Moving averages
Residual empirical process
Second order expansion

ABSTRACT

This paper discusses the problem of fitting a distribution function to the marginal distribution of a long memory moving average process. Because of the uniform reduction principle, unlike in the i.i.d. set up, classical tests based on empirical process are relatively easy to implement. More importantly, we discuss fitting the marginal distribution of the error process in location, scale, location–scale and linear regression models. An interesting observation is that in the location model, location–scale model, or more generally in the linear regression models with non-zero intercept parameter, the null weak limit of the first order difference between the residual empirical process and the null model is degenerate at zero, and hence it cannot be used to fit an error distribution in these models for the large samples. This finding is in sharp contrast to a recent claim of Chan and Ling (2008) that the null weak limit of such a process is a continuous Gaussian process. This note also proposes some tests based on the second order difference for the location case. Another finding is that residual empirical process tests in the scale problem are robust against not knowing the scale parameter.

1. Introduction

The problem of fitting a parametric family of distributions to a probability distribution, known as the goodness-of-fit testing problem, is well studied in the literature when the underlying observations are i.i.d. See, for example, Durbin (1973, 1975), Khmaladze (1979, 1981), D'Agostino and Stephens (1986), among others.

A discrete time stationary stochastic process with finite variance is said to have *long memory* if its autocorrelations tend to zero hyperbolically in the lag parameter, as the lag tends to infinity, but their sum diverges. The importance of these processes in econometrics, hydrology and other physical sciences is abundantly demonstrated in the works of Beran (1992, 1994), Baillie (1996), Dehling et al. (2002), and Doukhan et al. (2003), and the references therein.

We model long memory via moving averages. Let $\mathbb{Z} := \{0, \pm 1, \ldots\}$. We suppose for the time being that the observable process is

$$\varepsilon_j = \sum_{k=0}^{\infty} b_k \zeta_{j-k}, \quad j \in \mathbb{Z}, \tag{1.1}$$

* Corresponding author. Tel.: +370 5 2 729 207; fax: +370 5 2 729 209.
E-mail addresses: koul@stt.msu.edu (H.L. Koul), sdonatas@ktl.mii.lt (D. Surgailis).

[1] Research supported in part by the NSF DMS Grant 0704130.
[2] Research supported in part by the bilateral France-Lithuania scientific project Gilibert and the Lithuanian State Science and Studies Foundation Grant T-70/09.

0378-3758/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2010.04.039

where $\zeta_s, s \in \mathbb{Z}$, are i.i.d., with zero mean and unit variance. The constants b_j are assumed to satisfy $b_k = 0, k < 0, b_0 = 1$, and

$$b_j \sim c_0 j^{-(1-d)} \quad \text{as } j \to \infty \text{ for some } 0 < c_0 < \infty \text{ and } 0 < d < 1/2. \tag{1.2}$$

Let $B(a,b) := \int_0^1 x^{a-1}(1-x)^{b-1}\,dx, a>0, b>0$. One can directly verify that as $k \to \infty$,

$$\mathrm{Cov}(\varepsilon_0, \varepsilon_k) = \sum_{j=0}^{\infty} b_j b_{j+k} \sim c_0^2 \int_0^{\infty} y^{d-1}(1+y)^{d-1}\,dy\, k^{-(1-2d)} = c_0^2 B(d, 1-2d) k^{-(1-2d)},$$

so that the process $\varepsilon_j, j \in \mathbb{Z}$, has long memory.

Now, let ε denote a copy of ε_1, F denote the marginal d.f. of ε and F_0 be a known d.f. The problem of interest here is to test the simple hypothesis

$$H_0 : F = F_0 \quad \text{vs.} \quad H_1 : F \neq F_0.$$

This problem is of interest in applications. For example in the value at risk analysis, cf. Tsay (2002), various probability calculations are based on the assumption that the underlying process is a Gaussian process. If one would reject the null hypothesis of a marginal distribution being Gaussian then such analysis would be a suspect.

In this note we shall discuss asymptotic behavior of some omnibus tests for H_0 based on the empirical d.f. $F_n(x) := \sum_{i=1}^n I(\varepsilon_i \le x)/n$ of $\varepsilon_i, 1 \le i \le n$ for testing H_0.

A bit more interesting and at the same time surprisingly challenging problem is to test

$$H_0^{loc} : F(x) = F_0(x-\mu), \quad \forall x \in \mathbb{R} \text{ for some } \mu \in \mathbb{R}$$

vs.

$$H_1^{loc} : H_0^{loc} \text{ is not true.}$$

This problem is equivalent to testing for H_0 based on the observations

$$Y_i = \mu + \varepsilon_i, \quad i = 1, \ldots, n \text{ for some } \mu \in \mathbb{R}, \tag{1.3}$$

where now $\varepsilon_i, i \in \mathbb{Z}$, are unobservable moving average errors of (1.1). Here tests would be based on

$$\overline{F}_n(x) := n^{-1} \sum_{i=1}^n I(Y_i - \overline{Y}_n \le x), \quad \overline{Y}_n := n^{-1} \sum_{i=1}^n Y_i, \ x \in \mathbb{R}.$$

Sometimes we may be interested in testing the equivalence of F to F_0 up to a scale parameter, i.e., to test

$$H_0^{sc} : F(x) = F_0(x/\sigma), \quad \forall x \in \mathbb{R} \text{ for some } \sigma > 0$$

vs.

$$H_1^{sc} : H_0^{sc} \quad \text{is not true.}$$

This problem is equivalent to testing for H_0 based on the observations

$$Y_i = \sigma \varepsilon_i, \quad i = 1, \ldots, n \quad \text{for some } \sigma > 0, \tag{1.4}$$

where again $\varepsilon_i, i \in \mathbb{Z}$, are unobservable moving average errors of (1.1). Here tests will be based on

$$\tilde{F}_n(x) := n^{-1} \sum_{i=1}^n I(Y_i/\tilde{\sigma}_n \le x), \quad \tilde{\sigma}_n^2 := n^{-1} \sum_{i=1}^n Y_i^2, \ x \in \mathbb{R}.$$

Another interesting problem is to fit a distribution up to unknown location and scale parameters, i.e., to test the hypothesis

$$\mathcal{H}_0 : F(x) = F_0\left(\frac{x-\mu}{\sigma}\right), \quad \forall x \in \mathbb{R} \text{ and for some } \mu \in \mathbb{R}, \ \sigma > 0,$$

$$\mathcal{H}_1 : \mathcal{H}_0 \quad \text{is not true.}$$

This is equivalent to testing for H_0 based on the observations

$$Y_i = \mu + \sigma \varepsilon_i, \quad i = 1, \ldots, n \text{ for some } \mu \in \mathbb{R}, \sigma > 0, \tag{1.5}$$

where again $\varepsilon_i, i \in \mathbb{Z}$, are unobservable as in (1.1). Here tests will be based on

$$\mathcal{F}_n(x) := n^{-1} \sum_{i=1}^n I(Y_i - \overline{Y}_n \le x s_n), \quad s_n^2 := n^{-1} \sum_{i=1}^n (Y_i - \overline{Y}_n)^2, \ x \in \mathbb{R}.$$

Next, consider the problem of fitting marginal error d.f. in the linear regression model where one is given an array of $p \times 1$ design vectors $x_{ni}, 1 \le i \le n$, and one observes an array of random variables $\{Y_{ni}; 1 \le i \le n\}$ from the model

$$Y_{ni} = x_{ni}'\beta + \varepsilon_i, \quad 1 \le i \le n \text{ for some } \beta \in \mathbb{R}^p, \tag{1.6}$$

with $\varepsilon_i, i \in \mathbb{Z}$, as in (1.1). Now F denotes marginal d.f. of the error process. Consider the problem of testing the above H_0 vs. H_1 based on $Y_{ni}, 1 \le i \le n$. Let $\hat{\beta}_n$ be the LSE of β and $\hat{F}_n$ denote the empirical d.f. of the residuals $\hat{\varepsilon}_i := Y_{ni} - x_{ni}'\hat{\beta}_n, 1 \le i \le n$.

In the next section we discuss tests for the first problem based on F_n, while Section 3 pertains to tests for H_0^{loc}, H_0^{sc} and $\mathcal{H}_0$ based on $\overline{F}_n$, $\tilde{F}_n$, and $\mathcal{F}_n$, respectively. It is observed that the first order differences $n^{1/2-d}(\overline{F}_n-F_0)$ and $n^{1/2-d}(\mathcal{F}_n-F_0)$ cannot be used to test for H_0^{loc} and $\mathcal{H}_0$, respectively, while tests based on $n^{1/2-d}(\tilde{F}_n-F_0)$ for testing H_0^{sc} have the same large sample behavior as in the case of known σ.

Tests for fitting an error d.f. in linear regression set up based on $n^{1/2-d}(\hat{F}_n-F_0)$ are discussed in Section 4. This process also converges to zero in probability under H_0 if there is a non-zero intercept parameter in (1.6), and thus cannot be used to test for H_0 asymptotically. Section 5 suggests some tests of H_0^{loc} based on the second order difference for $\overline{F}_n-F_0$.

The recent paper of Chan and Ling (2008) discuss the first order asymptotics of residual empirical processes and goodness-of-fit tests for long-memory errors. They consider the regression models with random design. Although the first order approximations in their paper are similar to ours, some conclusions of Chan and Ling (2008) for hypotheses testing are incorrect (see Remark 4.1 below). Moreover, they do not discuss tests based on second order differences.

2. Tests for simple hypothesis

Here, we shall analyze asymptotic behavior of the first order process $n^{1/2-d}(F_n-F_0)$ when $\varepsilon_i, i \in \mathbb{Z}$, is an observable process. The Kolomorov test based on the supremum statistic

$$K_n := \sup_{x\in\mathbb{R}} |F_n(x)-F_0(x)|$$

will be discussed in some detail.

To proceed, we need to assume, for some $C<\infty$ and $\delta>0$,

$$|Ee^{iu\zeta_0}| \le C(1+|u|)^{-\delta}, \quad E|\zeta_0|^3<\infty. \tag{2.1}$$

Let $c(\theta)=c_0^2B(d,1-2d)/d(1+2d)$, $\theta := (c_0,d)'$, and $\overline{\varepsilon}_n := n^{-1}\sum_{i=1}^n \varepsilon_i$. Then, from Giraitis et al. (1996) (GKS) and Koul and Surgailis (2002), we obtain that under H_0, F_0 is infinitely differentiable with smooth and bounded Lebesgue density f_0, and

$$\sup_{x\in\mathbb{R}} |n^{1/2-d}(F_n(x)-F_0(x))+f_0(x)n^{1/2-d}\overline{\varepsilon}_n| = o_p(1) \quad (H_0). \tag{2.2}$$

Dehling and Taqqu (1989) proved an analog of (2.2) when $\varepsilon_i, i \in \mathbb{Z}$, is a stationary Gaussian process, and coined the phrase *uniform reduction principle* (URP) for this type of result. Koul and Mukherjee (1993) established URP when $\varepsilon_i, i \in \mathbb{Z}$, is subordinated to a Gaussian process, and GKS proved it when $\varepsilon_i, i \in \mathbb{Z}$, is a long memory moving average process and under a higher moment assumption on ζ_0. Koul and Surgailis (2002) obtained the above expansion under the condition (2.1) that includes the finite third moment assumption about ζ_0.

In the sequel, $u_p(1)$ denotes a sequence of stochastic processes indexed by $x \in \overline{\mathbb{R}}$ and tending to zero, uniformly over $x \in \overline{\mathbb{R}}$, in probability, and $\Longrightarrow(\to_D)$ denotes the weak convergence of a sequence of stochastic processes in the Skorohod space $D(\overline{\mathbb{R}})$ with the sup-topology (r.v.'s), where $\overline{\mathbb{R}} := [-\infty,\infty]$. For any smooth function g from $\mathbb{R}$ to $\mathbb{R}$, $\dot{g}$ and $\ddot{g}$ denote its first and second derivatives, respectively.

Now, by a result of Davydov (1970),

$$c^{-1/2}(\theta)n^{1/2-d}\overline{\varepsilon}_n \to_D \mathcal{N}(0,1). \tag{2.3}$$

Hence, from (2.2) we obtain

$$n^{1/2-d}(F_n(x)-F_0(x)) \implies -c^{1/2}(\theta)f_0(x)Z, \tag{2.4}$$

where $Z\sim\mathcal{N}(0,1)$ r.v. Consequently, with $\|f_0\|_\infty := \sup_{x\in\mathbb{R}} f_0(x)$,

$$n^{1/2-d}K_n \to_D c^{1/2}(\theta)|Z|\|f_0\|_\infty,$$

and we readily obtain that under H_0,

$$\mathcal{K}_n(\theta) := \frac{n^{1/2-d}K_n}{c^{1/2}(\theta)\|f_0\|_\infty} \to_D |Z|.$$

The expansion (2.2) and the limit distribution of K_n were also derived in Ho and Hsing (1996, (3.1), (3.2)), under more stringent assumptions on the distribution of the innovations ζ_j.

Implementation of the K_n test requires a consistent and a log(n)-consistent estimators of c_0 and d, respectively. Dalla et al. (2006) show that the semi-parametric local Whittle estimators $\hat{c}_0$ and $\hat{d}$ of c_0 and d satisfy these conditions. Let $\hat{\theta} := (\hat{c}_0,\hat{d})'$. The proposed test would reject H_0 whenever $\mathcal{K}_n(\hat{\theta})$ is large.

Clearly, the determination of the asymptotic critical values of this test is simple compared to its analog in the i.i.d. set up. For example, the test that rejects H_0 whenever $\mathcal{K}_n(\hat{\theta}) > z_{\alpha/2}$ would be of asymptotic size α, $0<\alpha<1$, where z_α is the $100(1-\alpha)$th percentiles of standard normal distribution.

Needless to say similar statements apply to any other tests based on continuous functionals of the first order difference $n^{1/2-d}(F_n-F_0)$. For example, the Cramér–von Mises test that rejects H_0 whenever $\mathcal{C}_n(\hat{\theta}) > \chi^2_{1-\alpha}$ would also be of asymptotic size α, where χ^2_α is the 100(1−α)th percentile of chi-square distribution with 1 degree of freedom. Here

$$C_n := \int [F_n(x) - F_0(x)]^2 dF_0(x), \quad \mathcal{C}_n(\theta) := \frac{n^{1-2d} C_n}{c(\theta) \int f_0^3(x)\, dx}.$$

These findings are thus in complete contrast to the results available in the i.i.d. set up where one must use the full knowledge of the distribution of Brownian bridge on [0,1] to implement tests based on the first order difference $n^{1/2}(F_n - F_0)$ for large samples, cf., Durbin (1973, 1975).

We shall now briefly analyze asymptotic power of the K_n test. Let $F \neq F_0$ be another marginal d.f. of ε such that (2.1) is satisfied and the local Whittle estimators $\hat{c}_0, \hat{d}$ are consistent and log(n)-consistent for c_0, d. Then, arguing as above, F has a smooth Lebesgue density f and

$$\mathcal{K}_n(\hat{\theta}) = \frac{\sup_{x \in \mathbb{R}} |-c^{1/2}(\theta) f(x) Z + n^{1/2-d}(F(x) - F_0(x))|}{c^{1/2}(\theta) \|f_0\|_\infty} + o_p(1).$$

From this we readily see that the above Kolmogorov test is consistent at this F. It has trivial asymptotic power against sequences of alternatives for which $\sup_x n^{1/2-d}|F(x) - F_0(x)| \to 0$. Thus this test cannot distinguish the $n^{1/2}$-neighborhoods of F_0, i.e., this test has asymptotic power α against those $\{F\}$ in the class of d.f. satisfying the above assumed conditions and for which $n^{1/2} \sup_x |F(x) - F_0(x)| = O(1)$. Similar conclusions would hold for other tests based on $n^{1/2-d}(F_n - F_0)$.

Asymptotic power of the $\mathcal{K}_n(\hat{\theta})$ test against the sequence of local alternatives $F = F_0 + n^{-(1/2-d)} \varDelta$, where $\varDelta$ is absolutely continuous with a.e. derivative $\dot{\varDelta}$ bounded, equals

$$P\left(\sup_x |-f_0(x) c^{1/2}(\theta) Z + \varDelta(x)| > z_{\alpha/2} c^{1/2}(\theta) \|f_0\|_\infty \right).$$

3. Testing for H_0^{loc}, H_0^{sc}, and $\mathscr{H}_0$

First, consider testing for H_0^{loc}. Let, now $\bar{\varepsilon}_n := \sum_{i=1}^n (Y_i - \mu)/n$. Recall that here $\overline{F}_n(x) = F_n(x + \bar{\varepsilon}_n)$.

Proposition 3.1 (*URP for the residual empirical process* $\overline{F}_n$.). *Assume* (1.2) *and* (2.1) *hold. Then,*

$$\sup_{x \in \mathbb{R}} n^{1/2-d} |\overline{F}_n(x) - F_0(x)| = o_p(1). \tag{3.1}$$

Proof. From (2.2), (2.4) and the mean value theorem one obtains

$$\begin{aligned} n^{1/2-d}(\overline{F}_n(x) - F_0(x)) &= n^{1/2-d} \left[F_n(x + \bar{\varepsilon}_n) - F_0(x + \bar{\varepsilon}_n) + \bar{\varepsilon}_n f_0(x + \bar{\varepsilon}_n) + \int_x^{x + \bar{\varepsilon}_n} (f_0(u) - f_0(x + \bar{\varepsilon}_n))\, du \right] \\ &= u_p(1) + O_p(\|\dot{f}_0\|_\infty n^{1/2-d} |\bar{\varepsilon}_n|^2) = u_p(1). \end{aligned}$$

This completes the proof. □

According to Proposition 3.1, the first order difference $D_n(x) := n^{1/2-d}(\overline{F}_n(x) - F_0(x))$ cannot distinguish between the two marginal distributions of a long memory moving average process that differ only in their means. This finding is in sharp contrast to what is available in the case of i.i.d. errors where the first order difference $n^{1/2}[\overline{F}_n(x) - F_0(x)]$ converges weakly to a time transformed Brownian bridge with a drift under H_0^{loc}, cf., Durbin (1973).

Next, consider the scale problem. Here, $\varepsilon_i = Y_i/\sigma$, and

$$\tilde{F}_n(x) = \frac{1}{n} \sum_{i=1}^n I(Y_i/\tilde{\sigma}_n \le x) = F_n(x \tilde{\sigma}_n/\sigma).$$

Proposition 3.2 (*URP for the residual empirical process* $\tilde{F}_n$). *Assume* (1.2), (2.1) *hold and* $E\zeta_0^4 < \infty$. *Then,*

$$\sup_{x \in \overline{\mathbb{R}}} |n^{1/2-d}[\tilde{F}_n(x) - F_0(x)] + n^{1/2-d} \bar{\varepsilon}_n f_0(x)| = o_p(1).$$

Proof. Without loss of generality, assume $E\varepsilon_i^2 = \sum_{j=0}^\infty b_j^2 = 1$. We have

$$E(\tilde{\sigma}_n^2 - \sigma^2)^2 = \sigma^4 n^{-2} E\left(\sum_{i=1}^n (\varepsilon_i^2 - 1) \right)^2 = \sigma^4 n^{-2} \sum_{i,j=1}^n \mathrm{Cov}(\varepsilon_i^2, \varepsilon_j^2).$$

Let $\chi_4 = E\zeta_0^4 + 3$ be the 4th cumulant of ζ_0. By elementary computation (see Giraitis and Surgailis, 1990, (2.3))

$$\mathrm{Cov}(\varepsilon_i^2,\varepsilon_j^2) = \left(\sum_{k=0}^{\infty} b_k b_{k+i-j}\right)^2 + \chi_4 \sum_{k=0}^{\infty} b_k^2 b_{k+i-j}^2 = (\mathrm{Cov}(\varepsilon_i,\varepsilon_j))^2 + \mathrm{O}(|i-j|^{-2(1-d)}),$$

implying $\mathrm{Cov}(\varepsilon_i^2,\varepsilon_j^2) = \mathrm{O}((\mathrm{Cov}(\varepsilon_i,\varepsilon_j))^2) = \mathrm{O}(|i-j|^{-2(1-2d)})$, as $|i-j| \to \infty$. Whence it follows that

$$\begin{aligned} \frac{\tilde\sigma_n}{\sigma} - 1 &= \mathrm{O_p}(n^{-1/2}), \quad 0<d<1/4 \\ &= \mathrm{O_p}((\log(n)/n)^{1/2}), \quad d=1/4 \\ &= \mathrm{O_p}(n^{-(1-2d)}), \quad 1/4<d<1/2. \end{aligned} \tag{3.2}$$

Now, let $\Delta_n := (\tilde\sigma_n/\sigma - 1)$. Recall from Lemma 5.1 of Koul and Surgailis (2002) (KS) that under (2.1), F_0 is infinitely smooth with density f_0 satisfying

$$\sup_{x\in\mathbb{R}} (1+x^2)(f_0(x) + |\dot f_0(x)|) < \infty.$$

This bound and (5.12) of Lemma 5.2 of KS imply

$$|F_0(x+x\Delta_n) - F_0(x)| = \left|\int_0^{x\Delta_n} f_0(x+u)\,du\right| \le C(1+x^2)^{-1}(|x\Delta_n| + x^2\Delta_n^2) \le C(|\Delta_n| + \Delta_n^2),$$

$$|f_0(x+x\Delta_n) - f_0(x)| = \left|\int_0^{x\Delta_n} \dot f_0(x+u)\,du\right| \le C(1+x^2)^{-1}(|x\Delta_n| + x^2\Delta_n^2) \le C(|\Delta_n| + \Delta_n^2). \tag{3.3}$$

Hence,

$$\begin{aligned} n^{1/2-d}|\tilde F_n(x) - F_0(x) + \bar\varepsilon_n f_0(x)| &\le n^{1/2-d}|F_n(x\tilde\sigma_n/\sigma) - F_0(x\tilde\sigma_n/\sigma) + \bar\varepsilon_n f_0(x\tilde\sigma_n/\sigma)| \\ &\quad + n^{1/2-d}\bar\varepsilon_n|f_0(x) - f_0(x+x\Delta_n)| + n^{1/2-d}|F_0(x+x\Delta_n) - F_0(x)| \le u_p(1) + Cn^{1/2-d}(|\Delta_n| + \Delta_n^2). \end{aligned}$$

The statement of the proposition now follows from this bound (3.2) and the fact that $0<d<\frac{1}{2}$. □

Note the URP for $\tilde F_n$ is precisely the same as for F_n given in (2.2). In other words, asymptotic null distribution of tests based on $n^{1/2-d}(\tilde F_n - F_0)$ for testing H_0^{sc} is the same as those of tests based on $n^{1/2-d}(F_n - F_0)$ for testing H_0. This is a kind of robustness property of these tests against the unknown error variance. It is also unlike the above situation in the location model, and unlike the situation in the i.i.d. set up, where $n^{1/2}(\tilde F_n - F_0)$ weakly converges to a Brownian bridge with a drift, cf., e.g., Durbin (1973).

Location–scale problem: Now consider the problem of testing for $\mathcal{H}_0$. Assume, as in the scale problem, that $E\zeta_0^4 < \infty$. Let $\delta_n := (s_n - \sigma)/\sigma$ and $\varepsilon_i = (Y_i - \mu)/\sigma$, $\bar\varepsilon_n := (\bar Y_n - \mu)/\sigma$. Then, with $\bar F_n$ the same as in Proposition 3.1,

$$\mathcal{F}_n(x) = n^{-1}\sum_{i=1}^n I(Y_i - \bar Y_n \le x s_n) = n^{-1}\sum_{i=1}^n I(\varepsilon_i \le x + x\delta_n + \bar\varepsilon_n) = \bar F_n(x + x\delta_n).$$

Also note that the bound (3.2) continues to hold with $\tilde\sigma_n$ replaced by s_n under the current set up. Using this fact, (3.1), and an argument like the one used for deriving the bound in (3.3), we readily obtain, under $\mathcal{H}_0$,

$$n^{1/2-d}\sup_x|\mathcal{F}_n(x) - F_0(x)| \le n^{1/2-d}\left\{\sup_x|\bar F_n(x) - F_0(x)| + \sup_x|F_0(x+x\delta_n) - F_0(x)|\right\} = o_p(1) + Cn^{1/2-d}(|\delta_n| + \delta_n^2) = o_p(1).$$

Thus, here also the location case dominates in the sense that the first order difference $n^{1/2-d}[\mathcal{F}_n(x) - F_0(x)], x \in \mathbb{R}$, is not useful for fitting a marginal d.f. up to the unknown location and scale parameters.

4. Fitting an error d.f. in a regression model

Recall the linear regression model (1.6) and definition of $\hat F_n$ from Section 1, viz.,

$$\hat F_n(x) = n^{-1}\sum_{i=1}^n I(Y_{ni} - x_{ni}'\hat\beta_n \le x) = n^{-1}\sum_{i=1}^n I(\varepsilon_i \le x + x_{ni}'(\hat\beta_n - \beta)),$$

where $\hat\beta_n$ is the LSE of $\beta \in \mathbb{R}^p$.

Proposition 4.1 (*URP for the residual empirical process* $\hat{F}_n$). *Assume the same conditions on errors as in Proposition* 3.1. *Assume the matrix* $\boldsymbol{X}_n$ *whose ith row is* $x'_{ni}, 1 \le i \le n$, *is of full rank. Let* $\boldsymbol{D}_n := (\boldsymbol{X}'_n \boldsymbol{X}_n)^{1/2}$. *Assume additionally*

$$n^{1/2} \max_{1 \le i \le n} \|\boldsymbol{D}_n^{-1} x_{ni}\| = O(1). \tag{4.1}$$

Then,

$$\sup_{x \in \overline{\mathbb{R}}} \left| n^{1/2-d}[\hat{F}_n(x) - F_0(x)] - f_0(x) n^{-1/2-d} \sum_{i=1}^n [x'_{ni}(\hat{\beta}_n - \beta) - \varepsilon_i] \right| = o_p(1). \tag{4.2}$$

Proof. To begin with we have uniform linearity of the residual empirical processes: For any *nonrandom* array $\{\xi_{ni}, 1 \le i \le n\}$ of real numbers such that $\max_{1 \le i \le n} |\xi_{ni}| = O(1)$,

$$\sup_{x \in \overline{\mathbb{R}}} \left| n^{-1/2-d} \sum_{i=1}^n [I(\varepsilon_i \le x + \xi_{ni} n^{d-1/2}) - I(\varepsilon_i \le x) - n^{d-1/2} \xi_{ni} f_0(x)] \right| = o_p(1). \tag{4.3}$$

This result was proved in GKS (Theorem 1(1.9)) under a higher moment assumption on ζ_0. Under the current third moment assumption it follows from (3.16) of Koul and Surgailis (2003) and an argument as in GKS.

The result (4.3) entails the following fact. For any $0 < b < \infty$ and any nonrandom array $\{c_{ni}, 1 \le i \le n, n \ge 1\}, c_{ni} \in \mathbb{R}^p$, with $\max_{1 \le i \le n} \|c_{ni}\| = O(1)$,

$$\sup_{x \in \overline{\mathbb{R}}, \|s\| \le b} n^{-1/2-d} \left| \sum_{i=1}^n [I(\varepsilon_i \le x + c'_{ni} s n^{d-1/2}) - I(\varepsilon_i \le x) - c'_{ni} s n^{d-1/2} f_0(x)] \right| = o_p(1). \tag{4.4}$$

This is proved using arguments as in Koul (2002). For the sake of completeness we are reproducing this argument in the Appendix below.

Let $v_{ni} := n^{1/2} \boldsymbol{D}_n^{-1} x_{ni}$. Now write $x'_{ni}(\hat{\beta}_n - \beta) = v'_{ni} n^{d-1/2} n^{-d} \boldsymbol{D}_n(\hat{\beta}_n - \beta)$. Recall that

$$\boldsymbol{D}_n(\hat{\beta}_n - \beta) = \boldsymbol{D}_n^{-1} \sum_{i=1}^n x_{ni} \varepsilon_i = n^{-1/2} \sum_{i=1}^n v_{ni} \varepsilon_i.$$

In view of (1.2), we have $|E\varepsilon_i \varepsilon_j| \le C|i-j|^{2d-1}$, for all $i \ne j$. Hence, in view of (4.1),

$$E\|\boldsymbol{D}_n(\hat{\beta}_n - \beta)\|^2 = n^{-1} \sum_{i,j=1}^n v'_{ni} v_{nj} E\varepsilon_i \varepsilon_j \le C n^{-1} \sum_{i,j=1}^n |E\varepsilon_i \varepsilon_j| \le C n^{2d},$$

$$n^{-d} \boldsymbol{D}_n(\hat{\beta}_n - \beta) = O_p(1).$$

Now (4.4) applied with $c_{ni} = v_{ni}$, $s = n^{-d} \boldsymbol{D}_n(\hat{\beta}_n - \beta)$ and a routine argument yields

$$\sup_{x \in \mathbb{R}} \left| n^{1/2-d}[\hat{F}_n(x) - F_n(x)] - n^{-d-1/2} \sum_{i=1}^n x'_{ni}(\hat{\beta}_n - \beta) f_0(x) \right| = o_p(1). \tag{4.5}$$

Write Δ_n for the l.h.s. of (4.2). From (4.5) and (2.2) we obtain

$$\Delta_n = \sup_{x \in \mathbb{R}} \left| n^{1/2-d}[\hat{F}_n(x) - F_n(x)] - f_0(x) n^{-d-1/2} \sum_{i=1}^n x'_{ni}(\hat{\beta}_n - \beta) + n^{1/2-d}[F_n(x) - F_0(x)] + f_0(x) n^{-d-1/2} \sum_{i=1}^n \varepsilon_i \right| = o_p(1),$$

proving (4.2). □

Now, let $\overline{v}_n = n^{-1} \sum_{i=1}^n v_{ni}$, and

$$\mathcal{Z}_n := n^{-1/2-d} \sum_{i=1}^n [x'_{ni}(\hat{\beta}_n - \beta) - \varepsilon_i] = n^{-1/2-d} \sum_{i=1}^n [\overline{v}'_n v_{ni} - 1] \varepsilon_i.$$

Remark 4.1. Note that if the regression model (1.6) is the one sample location model (1.3), i.e., if in (1.6), $p = 1, x_{ni} \equiv 1, \beta = \mu$, then $v_{ni} \equiv 1$, LSE is $\overline{Y}_n$, $\hat{F}_n(x) = \overline{F}_n(x)$ and $\mathcal{Z}_n = 0$, so that we again obtain the conclusion (3.1).

More generally, $\mathcal{Z}_n = 0$, for all $n \ge 1$, w.p.1, whenever there is a non-zero intercept parameter in the model (1.6). To see this and to keep the exposition transparent, consider the model (1.6) with $p=2$, $x'_{ni} = (1, a_i)$, where $a_1, \ldots, a_n$ are some known constants with $\sigma_a^2 := \sum_{i=1}^n (a_i - \overline{a})^2 > 0$, and $\overline{a} := n^{-1} \sum_{i=1}^n a_i$. Then,

$$\overline{v}'_n v_{ni} = n(1, \overline{a})(\boldsymbol{X}'_n \boldsymbol{X}_n)^{-1} \begin{pmatrix} 1 \\ a_i \end{pmatrix} = \frac{n}{\sigma_a^2}(1, \overline{a}) \begin{pmatrix} n^{-1} \sum_{j=1}^n a_j^2 & -\overline{a} \\ -\overline{a} & 1 \end{pmatrix} \begin{pmatrix} 1 \\ a_i \end{pmatrix} = 1, \quad \forall i = 1, \ldots, n.$$

Thus, as in the location problem, as long as there is a non-zero intercept parameter present in the regression model (1.6), the first order difference $n^{1/2-d}(\hat{F}_n - F_0)$ cannot be used to test for H_0 in these cases.

These facts thus contradict Corollary 3.1 of Chan and Ling (2008) which claims that $n^{1/2-d}(\hat{F}_n(x)-F_0(x)), x \in \mathbb{R}$, converges weakly to a continuous Gaussian process under the null hypothesis of the error d.f. being F_0. This corollary is incorrectly derived from the first order expansion in Chan and Ling (2008, Theorem 2.1) which is correct and agrees with the expansion in (4.5) above. The fact that they deal with random design does not change this contradiction.

Next, if in (1.6), $\sum_{i=1}^n x_{ni} = 0$, then $\mathcal{Z}_n = -n^{-1/2-d}\sum_{i=1}^n \varepsilon_i$, and by (2.3), we again obtain the analog of (2.4) for $\hat{F}_n$. In other words, if the design vectors corresponding to the slope parameters are orthogonal to the p vector of 1's, then asymptotic null distribution of $n^{1/2-d}(\hat{F}_n - F_0)$ is not affected by not knowing the slope parameters, and is the same as in (2.4). This fact has nothing to do with long memory. Analogous fact is available in the i.i.d. errors set up also, cf., Koul (2002, Chapter 6).

In general $\mathcal{Z}_n$ is a weighted sum of long memory moving average process. The following lemma gives a CLT for such r.v.'s.

Lemma 4.1. *Let $\varepsilon_i, i \in \mathbb{Z}$, be a linear process as in* (1.1) *and* (1.2). *Let $\{\mathbf{c}_{ni} \in \mathbb{R}^p, 1 \le i \le n\}$ be uniformly bounded nonrandom weights, i.e., $\sup_{n\ge 1}\sup_{1\le i\le n}|\mathbf{c}_{ni}| < \infty$ and let*

$$\mathbf{\Sigma}_n := \mathrm{Cov}\left(n^{-d-1/2}\sum_{i=1}^n \mathbf{c}_{ni}\varepsilon_i\right).$$

Assume that the limit $\lim_{n\to\infty}\mathbf{\Sigma}_n = \mathbf{\Sigma}$ exists and is positive definite. Then,

$$n^{-d-1/2}\sum_{i=1}^n \mathbf{c}_{ni}\varepsilon_i \to_D \mathcal{N}_p(0, \mathbf{\Sigma}).$$

The proof of Lemma 4.1 is given in Appendix. Now consider the weighted sum $\mathcal{Z}_n$. Clearly, under (4.1), the weights $c_{ni} \equiv \overline{v}_n' v_{ni} - 1$ are uniformly bounded. So one needs only the existence of the limit of $\mathbf{\Sigma}_n = \mathrm{Var}(\mathcal{Z}_n)$ in order to apply the above CLT to $\mathcal{Z}_n$.

As an example of design where this limit exists and the above lemma is applicable to $\mathcal{Z}_n$, consider the first degree polynomial regression through the origin where $p=1$ and $x_{ni} \equiv i/n$. Here, (4.1) is satisfied and $a_n := n^{1/2}D_n^{-1} = n^{1/2}(\sum_{i=1}^n (i/n)^2)^{-1/2} \sim 3^{1/2}$, $v_{ni} := (i/n)a_n$, $\overline{v}_n \sim 3^{1/2}/2$, and

$$\Sigma_n \equiv \Sigma_n(\theta) \sim 2c_0^2 B(d, 1-2d)\frac{1}{n^{1+2d}}\sum_{1\le i<j\le n}\left[\overline{v}_n\left(\frac{i}{n}\right)a_n - 1\right]\left[\overline{v}_n\left(\frac{j}{n}\right)a_n - 1\right]\frac{1}{(j-i)^{(1-2d)}}$$

$$\sim 2c_0^2 B(d, 1-2d)\left[\left(\frac{3}{2}\right)B(2, 2d) - \frac{1}{2d}\right]\left[\frac{3}{2(2+2d)} - \frac{1}{2d+1}\right] =: \Sigma(\theta) \equiv \Sigma.$$

Hence we can apply the above lemma to $\mathcal{Z}_n$ to conclude that $\mathcal{Z}_n \to_D \mathcal{N}(0, \Sigma(\theta))$. The role of $c(\theta)$ in the simple hypothesis case is now played by $\Sigma(\theta)$. Consequently, here the analog of Kolmogorov test for testing that the error d.f. is F_0 would be based on

$$\tilde{\mathcal{K}}_n := \frac{n^{1/2-d}\sup_x|\hat{F}_n(x) - F_0(x)|}{\Sigma(\tilde{\theta})^{1/2}\|f_0\|_\infty},$$

where now $\tilde{\theta}$ is the local Whittle estimator of θ based either on Y_{ni}' s or on the residuals $Y_{ni} - x_{ni}'\hat{\beta}_n$' s. Consistency of $\tilde{\theta}$ for θ under (4.1) follows from Dalla et al. (2006). Other tests based on $\hat{F}_n$ may be modified in a similar fashion. Needless to say these conclusions remain valid for any finite degree polynomial model.

5. Tests for H_0^{loc} based on the second order difference

In view of (3.1), it is desirable to ask the question whether there is a higher order difference of $\overline{F}_n - F_0$ that will provide a reasonable test for testing H_0^{loc}. In order to attempt to answer this question, we need to recall a result from Ho and Hsing (1996) and Koul and Surgailis (2002). Let $\varepsilon_j^{(0)} := 1$, $\varepsilon_j^{(1)} = \varepsilon_j$ of (1.1) and

$$\varepsilon_j^{(k)} := \sum_{s_k < \cdots < s_1 \le j} b_{j-s_1}\cdots b_{j-s_k}\zeta_{s_1}\cdots\zeta_{s_k},$$

be a polynomial of order $k \ge 2$ in the i.i.d. random variables $\zeta_s, s \in \mathbb{Z}$. This series converges in mean square for each $k \ge 1$, under the conditions $\sum_{k=0}^\infty b_k^2 < \infty, E\zeta_0^2 < \infty$ alone. Moreover, for each $k \ge 1$, the process $\varepsilon_j^{(k)}, j \in \mathbb{Z}$, is strictly stationary with zero mean and covariance

$$E(\varepsilon_0^{(k)}\varepsilon_j^{(k)}) = \sum_{0\le i_1<\cdots<i_k} b_{j+i_1}b_{i_1}\cdots b_{j+i_k}b_{i_k}, \tag{5.1}$$

and for any integers j, i, $E\varepsilon_j^{(k)}\varepsilon_i^{(\ell)}=0$ $(k\neq\ell,k,\ell=0,1,\ldots)$. It follows from (1.2) and (5.1) that for each $k\geq 1$,

$$|E(\varepsilon_0^{(k)}\varepsilon_j^{(k)})|\leq\frac{1}{k!}\left(\sum_{i=1}^{\infty}|b_{j+i}b_i|\right)^k=O(j^{-k(1-2d)}),\quad j\to\infty,$$

$$E\left(\sum_{j=1}^{n}\varepsilon_j^{(k)}\right)^2=O(n^{2-k(1-2d)}),\quad k(1-2d)<1,$$
$$=O(n\log(n)),\quad k(1-2d)=1,$$
$$=O(n),\quad k(1-2d)>1,\quad n\to\infty.$$

Let

$$Z_n^{(k)}:=n^{k(1/2-d)-1}\sum_{j=1}^{n}\varepsilon_j^{(k)},\quad k\geq 1.$$

Note that

$$Z_n^{(1)}=n^{1/2-d}\overline{\varepsilon}_n,\quad Z_n^{(2)}=n^{-2d}\sum_{j=1}^{n}\sum_{s_2<s_1\leq j}b_{j-s_1}b_{j-s_2}\zeta_{s_1}\zeta_{s_2}.\tag{5.2}$$

Assume $1/(1-2d)$ is not an integer and let k^* denote the greatest integer in $1/(1-2d)$, i.e., $k^*:=[1/(1-2d)]$. Introduce the multiple Wiener–Itô integral

$$Z^{(k)}:=\frac{c_0^k}{k!}\int_{\mathbb{R}^k}\left\{\int_0^1\prod_{j=1}^{k}(u-s_j)_+^{-(1-d)}du\right\}W(ds_1)\cdots W(ds_k)$$

w.r.t. a Gaussian white noise $W(ds)$, $E\,(W(ds))^2=ds$, which is well-defined for $1\leq k\leq k*$. From Surgailis (2003a) we obtain under the conditions (1.2) and $E\zeta_0^2<\infty$,

$$(Z_n^{(k)},0\leq k\leq k*)\to_D(Z^{(k)},0\leq k\leq k*).\tag{5.3}$$

Note that $Z^{(1)}$ is a Gaussian r.v. while $Z^{(2)}$ equals the Rosenblatt process at 1, cf. Taqqu (1975).

We are now ready to state the following theorem giving higher order expansion of empirical process due to Ho and Hsing (1996). See also KS.

Theorem 5.1. *Let* $\{\varepsilon_j\}$ *be a long memory moving average satisfying* (1.1) *and* (1.2). *Suppose the d.f. of* ζ_0 *is* k^*+3 *times differentiable with bounded, continuous and integrable derivatives, and* $E|\zeta_0|^4<\infty$. *Then, under* H_0,

$$F_n(x)-F_0(x)=\sum_{1\leq k\leq k^*}(-1)^k n^{-k(1-2d)/2}Z_n^{(k)}\frac{d^{k-1}f_0(x)}{dx^{k-1}}+n^{-1/2}Q_n(x),$$

with

$$\sup_{x\in\mathbb{R}}|Q_n(x)|=O_p(n^{\delta}),\quad\forall\delta>0.\tag{5.4}$$

As pointed out in KS, pp. 220–221, while one can show the weak convergence of all finite dimensional distributions of the remainder process $\{Q_n(x),x\in\overline{\mathbb{R}}\}$ to that of a continuous Gaussian process, proving tightness of this process remains an open technical problem. Because of this, in the case $0<d<\frac{1}{4}$ or $k^*=1$, the expansion (5.4) yields only the URP of (2.2) and is not useful in deriving the limiting distribution of the second order difference of F_n-F_0. But if $\{\varepsilon_j,j\in\mathbb{Z}\}$ is a long memory Gaussian process, Theorem 2.1 of KS has shown that the process $\{Q_n(x),x\in\overline{\mathbb{R}}\}$ converges weakly to a continuous Gaussian process. Below, we state this result for $k^*=1$, or $0<d<\frac{1}{4}$, only. Let

$$R_j(x):=I(\varepsilon_j\leq x)-F_0(x)+f_0(x)\varepsilon_j,\quad j\in\mathbb{Z},$$

$$Q_n(x):=n^{-1/2}\sum_{j=1}^{n}R_j(x)=n^{1/2}\{F_n(x)-F_0(x)+f_0(x)\overline{\varepsilon}_n\},\quad x\in\mathbb{R}.\tag{5.5}$$

Proposition 5.1. *Let* $\{\varepsilon_j\}$ *be a Gaussian long memory moving average satisfying* (1.1) *and* (1.2) *with* $0<d<\frac{1}{4}$. *Then,* $Q_n(x)\Longrightarrow Q(x)$, *where* $\{Q(x),x\in\mathbb{R}\}$ *is a continuous Gaussian process with zero mean and covariance function*

$$\mathrm{Cov}(Q(x),Q(y))=\sum_{j\in\mathbb{Z}}\mathrm{Cov}(R_0(x),R_j(y)).$$

This proposition and the expansion (5.4) yield the following facts.

Proposition 5.2. (i) *Assume the same conditions as in Theorem* 5.1, *and let* $\frac{1}{4}<d<\frac{1}{2}$. *Then*,

$$\sup_x |n^{1-2d}\{\overline{F}_n(x)-F_0(x)\}-\dot{f}_0(x)[Z_n^{(2)}-2^{-1}(Z_n^{(1)})^2]|=o_p(1), \tag{5.6}$$

and

$$n^{1-2d}\{\overline{F}_n(x)-F_0(x)\}\Longrightarrow \dot{f}_0(x)\mathcal{Y}, \tag{5.7}$$

where $\mathcal{Y}:=Z^{(2)}-2^{-1}(Z^{(1)})^2$. *Moreover*,

$$E\mathcal{Y}=-2^{-1}c(\theta),$$

$$E\mathcal{Y}^2=\frac{c_0^4B^2(d,1-2d)}{2d}\left\{\frac{1}{2(4d-1)}+\frac{1}{2d(1+2d)^2}-\frac{1}{d(4d+1)}-\frac{B(1+2d,1+2d)}{d}\right\}. \tag{5.8}$$

(ii) *If* $\{\varepsilon_j\}$ *is a Gaussian long memory moving average satisfying* (1.1) *and* (1.2) *with* $0<d<\frac{1}{4}$, *then*

$$n^{1/2}\{\overline{F}_n(x)-F_0(x)\}\Longrightarrow Q(x), \tag{5.9}$$

where $Q(x),x\in\overline{\mathbb{R}}$ *is the Gaussian process of Proposition* 5.1.

Proof. (i) Note that $d>\frac{1}{4}$ implies $2d-\frac{1}{2}>0$ and $k^*\geq 2$. Combine these facts with (5.4) and (5.6) with a $\delta<2d-\frac{1}{2}$ to obtain the second order expansion under H_0:

$$\sup_{x\in\overline{\mathbb{R}}}|n^{1-2d}\{F_n(x)-F_0(x)+f_0(x)\overline{\varepsilon}_n\}-\dot{f}_0(x)Z_n^{(2)}|=o_p(1). \tag{5.10}$$

The decomposition

$$\overline{F}_n(x)-F_0(x)=F_n(x+\overline{\varepsilon}_n)-F_0(x+\overline{\varepsilon}_n)+F_0(x+\overline{\varepsilon}_n)-F_0(x)$$

and (5.10) yield

$$\sup_{x\in\mathbb{R}}|n^{1-2d}\{\overline{F}_n(x)-F_0(x)+\overline{\varepsilon}_nf_0(x+\overline{\varepsilon}_n)\}-\dot{f}_0(x+\overline{\varepsilon}_n)Z_n^{(2)}-n^{1-2d}\{F_0(x+\overline{\varepsilon}_n)-F_0(x)\}|=o_p(1).$$

Using Taylor expansion of $f_0,\dot{f}_0$ and the boundedness of $\ddot{f}_0$, we thus obtain

$$\begin{aligned}
n^{1-2d}\{\overline{F}_n(x)-F_0(x)\}&=-n^{1-2d}\overline{\varepsilon}_nf_0(x+\overline{\varepsilon}_n)+\dot{f}_0(x+\overline{\varepsilon}_n)Z_n^{(2)}-n^{1-2d}\{F_0(x+\overline{\varepsilon}_n)-F_0(x)\}+u_p(1)\\
&=-n^{1-2d}\overline{\varepsilon}_n\left[f_0(x)+\overline{\varepsilon}_n\dot{f}_0(x)+\frac{(\overline{\varepsilon}_n)^2}{2}\ddot{f}_0(x+\xi_n)\right]+[\dot{f}_0(x)+\overline{\varepsilon}_n\ddot{f}_0(x+\xi_n)]Z_n^{(2)}\\
&\quad+n^{1-2d}\left[\overline{\varepsilon}_nf_0(x)+\frac{(\overline{\varepsilon}_n)^2}{2}\dot{f}_0(x)+\frac{(\overline{\varepsilon}_n)^3}{6}\ddot{f}_0(x+\xi_n)\right]+u_p(1)\\
&=\dot{f}_0(x)\left[Z_n^{(2)}-\frac{(\overline{\varepsilon}_n)^2}{2}n^{1-2d}\right]+O_p(n^{1-2d}(\overline{\varepsilon}_n)^3)+u_p(1),
\end{aligned}$$

where ξ_n is a sequence of r.v.'s with $|\xi_n|\leq|\overline{\varepsilon}_n|$, $(\overline{\varepsilon}_n)^2n^{1-2d}=(Z_n^{(1)})^2\rightarrow_D(Z^{(1)})^2$ and $n^{1-2d}(\overline{\varepsilon}_n)^3=o_p(n^{d-1/2})=o_p(1)$. This proves (5.6). Claim (5.7) follows from (5.6) and (5.3).

Next, we prove (5.8). From definition (5.3) and the diagram formula of Wiener–Itô integrals (Surgailis, 2003b) one obtains $EZ^{(2)}$=0,

$$E(Z^{(1)})^2=c_0^2\int_{\mathbb{R}}\left\{\int_0^1(u-s)_+^{d-1}\,du\right\}^2ds=\frac{c_0^2B(d,1-2d)}{d(1+2d)}=c(\theta),$$

$$E(Z^{(1)})^4=3(E(Z^{(1)})^2)^2=3c^2(\theta),$$

$$E(Z^{(2)})^2=2^{-1}c_0^4\int_{\mathbb{R}^2}\left\{\int_0^1(u-s_1)_+^{d-1}(u-s_2)_+^{d-1}\,du\right\}^2ds_1\,ds_2=\frac{c_0^4B^2(d,1-2d)}{4d(4d-1)},$$

and

$$\begin{aligned}
EZ^{(2)}(Z^{(1)})^2&=c_0^4\int_{\mathbb{R}}\int_{\mathbb{R}}ds_1\,ds_2\int_0^1(u-s_1)_+^{d-1}(u-s_2)_+^{d-1}\,du\int_0^1(v-s_1)_+^{d-1}dv\int_0^1(w-s_2)_+^{d-1}\,dw\\
&=c_0^4B^2(d,1-2d)\int_0^1\int_0^1\int_0^1|u-v|^{2d-1}|u-w|^{2d-1}\,du\,dv\,dw\\
&=c_0^4B^2(d,1-2d)\int_0^1\{(2d)^{-1}(u^{2d}+(1-u)^{2d})\}^2\,du=\frac{c_0^4B^2(d,1-2d)}{2d^2}\left(\frac{1}{4d+1}+B(2d+1,2d+1)\right). \qquad\square
\end{aligned}$$

Proof of (ii). Using the definition of $Q_n(x)$ in (5.5), rewrite $n^{1/2}(\overline{F}_n(x)-F_0(x)) = Q_n(x+\overline{\varepsilon}_n)+U_n(x)$, where $U_n(x) := n^{1/2}[F_0(x+\overline{\varepsilon}_n)-F_0(x)-f_0(x+\overline{\varepsilon}_n)\overline{\varepsilon}_n]$. By Taylor expansion,

$$\sup_x |U_n(x)| \le n^{1/2}(\overline{\varepsilon}_n)^2 \sup_x |\dot{f}_0(x)| = O_p(n^{2d-1/2}) = o_p(1).$$

Therefore (5.9) follows from (5.5) and

$$\sup_x |Q_n(x+\overline{\varepsilon}_n)-Q_n(x)| = o_p(1). \tag{5.11}$$

For any $\delta_1,\delta_2 > 0$,

$$P\left(\sup_x |Q_n(x+\overline{\varepsilon}_n)-Q_n(x)| > \delta_1\right) \le P(|\overline{\varepsilon}_n| \ge \delta_2)+P\left(\sup_{|x-y|<\delta_2} |Q_n(x)-Q_n(y)| > \delta_1\right).$$

Since the sequence $Q_n, n \ge 1$, is tight in the sup-topology on $\overline{\mathbb{R}}$, the last probability can be made arbitrarily small for all n large enough, by choosing δ_2 small enough. As $\overline{\varepsilon}_n = o_p(1)$, this proves (5.11) and, also completes the proof of Proposition 5.2. □

Now, let $\hat{d}$ be a $\log(n)$ consistent estimator of d and $\mathcal{D}_n := \sup_{x\in\mathbb{R}} n^{1-2\hat{d}}|\overline{F}_n(x)-F_0(x)|$. Proposition 5.2(i) readily implies that in the case $\frac{1}{4} < d < \frac{1}{2}$, the asymptotic null distribution of $\mathcal{D}_n$ is the same as that of $\|\dot{f}_0\|_\infty \mathcal{Y}$. Consequently, asymptotic critical values of the test that rejects H_0^{loc} whenever $\mathcal{D}_n/\|\dot{f}_0\|_\infty$ is large can be determined from the distribution of $\mathcal{Y}$. Unfortunately this distribution depends on c_0, d in a complicated fashion, and is not easy to track. However, one may use a Monte Carlo method to simulate the distribution of this r.v. corresponding to $c_0 = \hat{c}_0, d = \hat{d}$.

Higher order moments of the limiting r.v. $\mathcal{Y}$ can be computed using the diagram formula for Wiener–Itô integrals; see e.g., Major (1981, Corollary 5.4) and Surgailis (2003b, Proposition 5.1), although the resulting formulas are rather cumbersome. The characteristic function of r.v. $Z^{(2)}$ and the representation of it as a infinite weighted sum of chi-squares was obtained by Rosenblatt (1961). See also Taqqu (1975, (6.2)).

For the case $0 < d < \frac{1}{4}$ and when the process ε_j is Gaussian satisfying (1.1) and (1.2), Proposition 5.2(ii) implies that the large sample critical values of the test that rejects H_0^{loc} whenever $\sup_{x\in\mathbb{R}} n^{1/2}|\overline{F}_n(x)-F_0(x)|$ is large can be determined from the distribution of $\|Q\|_\infty$. This is similar to the 'usual' asymptotics of the Kolmogorov test for short memory errors based on F_n, see e.g., Csörgő and Mielniczuk (1996).

Acknowledgements

The authors are grateful to the two referees for valuable comments and careful reading of the manuscript.

Appendix A

Proof of (4.4). For $s \in \mathbb{R}^p, z, y \in \mathbb{R}$ put

$$W(s,z;y) := n^{-1/2-d} \sum_{i=1}^n [I(\varepsilon_i \le y+(c_{ni}'s+\|c_{ni}\|z)n^{d-1/2}) - I(\varepsilon_i \le y) - f_0(y)(c_{ni}'s+\|c_{ni}\|z)n^{d-1/2}].$$

Fact (4.3) applied with $\xi_{ni} = c_{ni}'s+\|c_{ni}\|z$ implies

$$\sup_y |W(s,z;y)| = o_p(1), \quad \forall s \in \mathbb{R}^p,\ z \in \mathbb{R}. \tag{A.1}$$

Note $W(s;y) := W(s,0;y)$ is the empirical process on the left hand side of (4.4). It suffices to show that for any $\varepsilon > 0$ there exists an $N_\varepsilon \ge 1$ such that for any $n > N_\varepsilon$ and for every $0 < b < \infty$,

$$P\left(\sup_{y\in\mathbb{R}, \|t\| \le b} |W(t;y)| > 2\varepsilon\right) < 2\varepsilon. \tag{A.2}$$

Fix a $0 < b < \infty$ and an $\varepsilon > 0$. Because the ball $\mathcal{K}_b := \{s \in \mathbb{R}^p; \|s\| \le b\}$ is compact, for any $\delta > 0$ there exist $k = k(\delta, b)$ and points $s_1, \ldots, s_k$ in $\mathcal{K}_b$ such that for any $t \in \mathcal{K}_b$, $\|t-s_j\| < \delta$, for some $j=1,2,\ldots,k$. Hence,

$$P\left(\sup_{y\in\mathbb{R}, \|t\| \le b} |W(t;y)| > 2\varepsilon\right) \le \sum_{j=1}^k P\left(\sup_{t\in\mathcal{K}_b, \|t-s_j\|<\delta, y\in\mathbb{R}} |W(t;y)-W(s_j;y)| > \varepsilon\right) + \sum_{j=1}^k P\left(\sup_y |W(s_j;y)| > \varepsilon\right). \tag{A.3}$$

Note that for any $t \in \mathbb{R}^p$ such that $\|t-s_j\| < \delta$,

$$c_{ni}'s_j - \|c_{ni}\|\delta \le c_{ni}'t \le c_{ni}'s_j + \|c_{ni}\|\delta, \quad \forall 1 \le i \le n.$$

Let $\delta_1 := \varepsilon/(18C\|f\|_\infty)$, where C is such that $\max_i\|c_{ni}\| \le C$. The above inequality and the monotonicity of the indicator function imply that for all $\delta < \delta_1$, and for all $\|t-s_j\| < \delta$,

$$|W(t;y)-W(s_j;y)| \le |W(s_j,\delta_1;y)|+|W(s_j,-\delta_1;y)|+n^{-1}\sum_{i=1}^n \|c_{ni}\|f_0(y)(2\delta_1+\delta) \le |W(s_j,\delta_1;y)|+|W(s_j,-\delta_1;y)|+3Cf_0(y)\delta_1.$$

Introduce the events

$$\mathcal{A}_{nj} := \left\{ \sup_{\|t-s_j\| < \delta; y\in\mathbb{R}} |W(t;y)-W(s_j;y)| > \varepsilon \right\},$$

$$\mathcal{B}_{nj} := \left\{ \sup_y |W(s_j,\delta_1;y)| + \sup_y |W(s_j,-\delta_1;y)| > \varepsilon/2 \right\}.$$

Then, by (A.1), applied with $z = \pm\delta_1$, there is an $N_{j,\varepsilon}$ such that for all $n > N_{j,\varepsilon}$,

$$P(\mathcal{B}_{nj}) \le \varepsilon/k \quad \text{for each } j=1,\ldots,k.$$

Let $\mathcal{B}_{nj}^c$ denote the complement of $\mathcal{B}_{nj}$. On $\mathcal{B}_{nj}^c$,

$$\sup_{\|t-s_j\| < \delta; y\in\mathbb{R}} |W(t;y)-W(s_j;y)| \le \varepsilon/2+3C\|f_0\|_\infty\delta_1 = \varepsilon/2+\varepsilon/6 = 2\varepsilon/3 < \varepsilon,$$

for all j=1,…,k. Hence, for all $n > N_\varepsilon := \max_{1\le j\le k} N_{j,\varepsilon}$,

$$P(\mathcal{A}_{nj}) = P(\mathcal{A}_{nj}\cap\mathcal{B}_{nj}) + P(\mathcal{A}_{nj}\cap\mathcal{B}_{nj}^c) \le \varepsilon/k, \quad j=1,\ldots,k.$$

Similarly as above, by (A.1) applied with z=0, there is an M_ε such that for all $n > M_\varepsilon$,

$$P\left(\sup_y |W(s_j;y)| > \varepsilon\right) < \varepsilon/k, \quad j=1,\ldots,k.$$

Hence, for all $n > N_\varepsilon \vee M_\varepsilon$,

$$P\left(\sup_{y\in\mathbb{R},\|t\|\le b} |W(t;y)| > 2\varepsilon\right) \le \sum_{j=1}^k P(\mathcal{A}_{nj}) + \sum_{j=1}^k P\left(\sup_y |W(s_j;y)| > \varepsilon\right) < 2\varepsilon,$$

proving (A.2) and hence, (4.4). □

Proof of Lemma 4.1. Recall $E\zeta_0^2 = 1$. For a given $\mathbf{a} \in \mathbb{R}^p$, introduce a normal r.v. $U \sim \mathcal{N}(0,\mathbf{a}'\mathbf{\Sigma}\mathbf{a})$. In (1.2), define $b_j = 0, j < 0$, and let

$$b_{nj} := n^{-d-1/2}\sum_{i=1}^n \mathbf{a}'\mathbf{c}_{ni}b_{i-j}, \quad U_n := n^{-d-1/2}\sum_{i=1}^n \mathbf{a}'\mathbf{c}_{ni}\varepsilon_i = \sum_{j\in\mathbb{Z}} b_{nj}\zeta_j. \tag{A.4}$$

Then, $\mathrm{Var}(U_n) = \mathbf{a}'\mathbf{\Sigma}_n\mathbf{a} \to \mathbf{a}'\mathbf{\Sigma}\mathbf{a} = \mathrm{Var}(U)$ and the statement of the lemma translates to $U_n \to_D U$.

Introduce truncated i.i.d. r.v.'s $\zeta_{i,K} := \zeta_i I(|\zeta_i| \le K) - E\zeta_i I(|\zeta_i| \le K)$ $(i \in \mathbb{Z})$, and let $U_{n,K}$ be defined as in (A.4) with ζ_j' s replaced by $\zeta_{j,K}$' s. Then $E(U_n-U_{n,K})^2 = \mathrm{Var}(\zeta_0-\zeta_{0,K})\mathrm{Var}(U_n) \to 0$, as $K \to \infty$, uniformly in $n \ge 1$. Therefore, it suffices to show the asymptotic normality of $U_{n,K}$ instead of U_n, for each K fixed. Since $\zeta_{i,K}$ are bounded r.v.'s, it thus suffices to prove the claimed result under the assumption that the r.v. ζ_i' s are bounded having all moments finite. Then $U_n \to_D U$ follows if we show that

$$\mathrm{Cum}_k(U_n) = o(1), \quad \forall k = 3,4,\ldots, \tag{A.5}$$

where $\mathrm{Cum}_k(Y)$ denotes the kth order cumulant of the r.v. Y. Let $\chi_k := \mathrm{Cum}_k(\zeta_0)$. According to the multilinearity and additivity properties of cumulants,

$$\mathrm{Cum}_k(U_n) = \chi_k\sum_{j\in\mathbb{Z}} b_{nj}^k \le |\chi_k|\sup_{j\in\mathbb{Z}}|b_{nj}|^{k-2}\mathrm{Var}(U_n)$$

since $\sum_{j\in\mathbb{Z}} b_{nj}^2 = \mathrm{Var}(U_n)$, see (A.4). Therefore, (A.5) follows from

$$\sup_{j\in\mathbb{Z}}|b_{nj}|^{k-2} = o(1), \quad \forall k = 3,4,\ldots\ . \tag{A.6}$$

But,

$$|b_{nj}| \le Cn^{-d-1/2}\sum_{i=1}^n (i-j)_+^{d-1} = O(n^{-1/2}) = o(1), \quad \forall j \in \mathbb{Z}.$$

This proves (A.6) and also Lemma 4.1. □

References

Baillie, R.T., 1996. Long memory processes and fractional integration in econometrics. J. Econometrics 73, 5–59.
Beran, J., 1992. Statistical methods for data with long-range dependence (with discussion). Statist. Sci. 7, 404–427.
Beran, J., 1994. Statistics for long memory processes. Monographs on Statistics and Applied Probability, vol. 61. Chapman & Hall, New York.
Chan, N.G., Ling, S., 2008. Residual empirical processes for long and short memory time series. Ann. Statist. 36, 2453–2470.
Csörgő, S., Mielniczuk, J., 1996. The empirical process of a short-range dependent stationary sequence under Gaussian subordination. Probab. Theory Relat. Fields 104, 15–25.
D'Agostino, R.B., Stephens, M. (Eds.), 1986. Goodness-of-fit techniques. Statistics: Text-books and Monographs, vol. 68. Marcel Dekker, New York.
Dalla, V., Giraitis, L., Hidalgo, J., 2006. Consistent estimation of the memory parameter for nonlinear time series. J. Time Ser. Anal. 27, 211–251.
Davydov, Y., 1970. The invariance principle for stationary process. Theory Probab. Appl. 15, 145–180.
Dehling, H., Taqqu, M.S., 1989. The empirical process of some long-range dependent sequences with an application to U-statistics. Ann. Statist. 17, 1767–1783.
Dehling, H., Mikosch, T., Sørensen, M., 2002. Empirical Process Techniques for Dependent Data. Birkhäuser, Boston.
Doukhan, P., Oppenheim, G., Taqqu, M.S., 2003. Theory and Applications of Long-Range Dependence. Birkhäuser, Boston.
Durbin, J., 1973. Distribution Theory for Test based on the Sample d.f.. SIAM, Philadelphia.
Durbin, J., 1975. Kolmogorov–Smirnov tests when parameters are estimated with applications to tests of exponentiality and tests on spacings. Biometrika 62, 5–22.
Giraitis, L., Surgailis, D., 1990. A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittle's estimate. Prob. Theory Relat. Fields 86, 87–104.
Giraitis, L., Koul, H.L., Surgailis, D., 1996. Asymptotic normality of regression estimators with long memory errors. Statist. Probab. Lett. 29, 317–335.
Ho, H.-C., Hsing, T., 1996. On the asymptotic expansion of the empirical process of long memory moving averages. Ann. Statist. 24, 992–1024.
Khmaladze, E.V., 1979. The use of ω^2 tests for testing parametric hypotheses. Theory Probab. Appl. 24, 283–301.
Khmaladze, E.V., 1981. Martingale approach in the theory of goodness-of-fit tests. Theory Probab. Appl. 26, 240–257.
Koul, H.L., 2002. Weighted Empirical Processes in Dynamic Nonlinear Models. Lecture Notes in Statistics, vol. 166, second ed. Springer, New York.
Koul, H.L., Mukherjee, K., 1993. Asymptotics of R-, MD- and LAD-estimators in linear regression models with long range dependent errors. Probab. Theory Relat. Fields 95, 535–553.
Koul, H.L., Surgailis, D., 2002. Asymptotic expansion of the empirical process of long memory moving averages. In: Dehling, H., Mikosch, T., Sørensen, M. (Eds.), Empirical Process Techniques for Dependent Data. Birkhäuser, Boston, pp. 213–239.
Koul, H.L., Surgailis, D., 2003. Robust estimators in regression models with long memory errors. In: Doukhan, P., Oppenheim, G., Taqqu, M.S. (Eds.), Theory and Applications of Long Range Dependence. Birkhäuser, Boston, pp. 339–353.
Major, P., 1981. Multiple Wiener–Itô Integrals. Lecture Notes in Mathematics, vol. 849. Springer, Berlin.
Rosenblatt, M., 1961. Independence and dependence. In: Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, Berkeley, pp. 411–443.
Surgailis, D., 2003a. Non-CLTs U-statistics, multinomial formula and approximations of multiple Itô–Wiener integrals. In: Doukhan, P., Oppenheim, G., Taqqu, M.S. (Eds.), Theory and Applications of Long-Range Dependence. Birkhäuser, Boston, pp. 129–142.
Surgailis, D., 2003b. CLTs for polynomials of linear sequences: diagram formula with illustrations. In: Doukhan, P., Oppenheim, G., Taqqu, M.S. (Eds.), Theory and Applications of Long-Range Dependence. Birkhäuser, Boston, pp. 111–127.
Taqqu, M.S., 1975. Weak convergence to fractional Brownian motion and to the Rosenblatt process. Z. Wahrsch. Verw. Gebiete 31, 287–302.
Tsay, R.S., 2002. Analysis of Financial Time Series. In: Wiley Series in Probability and Statistics, New York.