Asymptotic theory for regression models with fractional local to unity root errors

Kris De Brabanter\textsuperscript{1,2} · Farzad Sabzikar\textsuperscript{1}

Received: 23 March 2020 / Accepted: 13 February 2021 / Published online: 11 March 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract
This paper develops the asymptotic theory for parametric and nonparametric regression models when the errors have a fractional local to unity root (FLUR) model structure. FLUR models are stationary time series with semi-long range dependence property in the sense that their covariance function resembles that of a long memory model for moderate lags but eventually diminishes exponentially fast according to the presence of a decay factor governed by an exponential tempering parameter. When this parameter is sample size dependent, the asymptotic theory for these regression models admit a wide range of stochastic processes with behavior that includes long, semi-long, and short memory processes.

Keywords Tempered linear processes · Semi-long range dependence · Non-parametric regression · Piecewise polynomial regression · Tempered fractional calculus

Mathematics Subject Classification 60G22 · 60G50 · 62F12 · 62G08

1 Introduction

The local to unity root (LUR) model time series \( \{X(t)\} \) is generated by

\[
X(t) = \rho_N X(t - 1) + \zeta(t), \quad t = 1, \ldots, N; \ X(0) = 0,
\]

(1)
where $\rho_N = 1 - \zeta$ for $c > 0$, $\{\xi(j)\}_{j \in \mathbb{Z}}$ are i.i.d innovations, and $N$ is the sample size (Bobkoski 2011; Chan and Wei 1987; Phillips 1987). The LUR models have been proven useful for analyzing economic models that can capture characteristics such as bubbles of various financial time series during the subprime crisis (Phillips and Yu 2011; Phillips et al. 2015a, b). Using the backward shift operator $B X(t) = X(t - 1)$, the LUR model $\{X(t)\}$ (1) can be extended to

$$
X_{d, \rho_N}(t) = (1 - \rho_N B)^{-d} \xi(t) = \sum_{k=0}^{\infty} \rho_N^k \omega_{-d}(k) \xi(t - k), \quad t \in \mathbb{Z},
$$

where $\omega_{-d}(k) = \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)}$ and $d \in \mathbb{R} \setminus \mathbb{N}_-$. We observe that $\rho^k_N \to 1$ as $N \to \infty$ and hence $X_{d, 1}$ forms a fractional time series including the well known ARFIMA model with long-range dependence (or long memory) when $0 < d < 1/2$. Fractionally integrated (FI) time series with long memory property have proven useful in many disciplines, including economics and finance (Granger and Joyeux 1980; Hosking 1981; Baillie 1996) and hydrology (Hosking 1984). However, in practice, one of the challenges of the long memory model is capturing the low frequency variability of data series due to the fact that the spectral density of the model has a pole at zero frequency. This is an important issue since the shape of spectrum near frequency zero plays a crucial role in long run prediction of time series (Müller and Watson 2014). The existence of $\rho_N$ in (2) makes the FLUR model an appropriate model that can be applied to better capture the low frequencies than the fractional time series with long memory for real data in many research areas such as finance, geophysics, turbulence, and climate change. For example (i) Fig. 1a shows spectral density for a North American Regional Climate Change Assessment Program (NARCCAP) climate data (circles), along with fitted fractional long memory spectrum (dashed) with memory parameter $d = 0.933$. The spectrum of the fractional long memory shows a lack of fit at low frequencies. More importantly, the fractional long memory model with $d = 0.933$ is not stationary while the data is stationary. From this figure, we can see that using the FLUR model (with $d = 0.9333$, $c = 2752.1$ and $N = 10, 585$) one can track the low frequency more accurately compared to fractional long memory. (ii) The adjusted closing price $C_t$ for General Electric (GE) from Yahoo Finance from 7/3/2000 to 7/1/2019 ($N = 4779$) was used to compute log returns $R_t = \ln(C_t/C_{t-1})$, which appear uncorrelated. The fitted parameters are $d = 0.998$ and $c = 14.337$. Fig. 1b shows that the spectral density of FLUR model provides a reasonable fit to the periodogram, which follows a power law at moderate frequencies, but levels off at low frequencies.

Sabzikar and Surgailis (2018b) applied the idea of tempered fractional calculus to introduce the tempered linear process (TLP) $\{X_{d, \lambda}(j)\}_{j \in \mathbb{Z}}$ with moving averages

$$
X_{d, \lambda}(j) = \sum_{k=0}^{\infty} e^{-\lambda k} b_d(k) \xi(j - k), \quad j \in \mathbb{Z},
$$
where \( \{\xi(j)\}_{j \in \mathbb{Z}} \) are i.i.d innovations, \( b_d(k) \) regularly varying at infinity as \( k^{d-1} \), viz.

\[
b_d(k) \sim \frac{c_0}{\Gamma(d)} k^{d-1}, \quad k \to \infty, \quad c_0 \neq 0, \quad d \neq 0,
\]

where \( d \in \mathbb{R} \setminus \mathbb{N} \) and \( \lambda > 0 \) is the tempering parameter. Since \( \rho_N \sim e^{-\lambda N} \) as \( N \to \infty \), we can replace \( \rho_N^d \) with \( e^{-\lambda N k} \) in (2) and interpret the FLUR model \( X_{d,\rho_N} \) in (2) as the tempered linear process \( X_{d,\lambda N} \).

The autocovariogram of FLUR and TLP resemble long range dependent series out to moderate lag lengths but eventually decays exponentially fast. Giraitis et al. (2000) named this behavior semi-long memory which is analogous to the semi-heavy tail property in Barndorff-Nielsen (1998). Giraitis et al. (2003) introduced semi-long memory ARFIMA(0, d, 0), semi-long memory LARCH, and semi-long memory ARCH processes and used these models to investigate the power and robustness of the \( R/S \) type tests under contiguous and semi-long memory alternatives. Dacorogna et al. (1993) and Granger and Ding (1996) argued that the covariance function of some economic time series decay slowly at first but ultimately decay much faster, such as the magnitude of certain powers of financial returns.

On the other hand, asymptotic theory for parametric and nonparametric regression models has proved useful in many empirical applications; see for example the vast literature when errors in these regression models exhibit a long memory model (Csörgő and Mielniczuk 1995a, b, c; Robinson 1997; Deo 1997; Guo and Koul 2007). As a result, the asymptotic results involve fractional Brownian motion (FBM) with an unknown parameter that has to be estimated from data to facilitate inference. Also, it is known that FBM is not a semi-martingale when the memory parameter \( d \neq 0 \). Therefore, we cannot use Itô stochastic calculus and, instead, we have to rely on Malliavin calculus which is significantly more complex.

Motivated by the aforementioned applications and features of the FLUR model and some issues with long memory models in regression theory (as described in the previous paragraph), we develop the asymptotic theory for parametric and nonparametric regression models when the errors follow a FLUR model. Next, we explain the main contributions of this work which will be discussed in Sects. 3 and 4. In Sect. 3, we
consider the nonparametric regression model in fixed-design

\[ Y(j) = m\left(\frac{j}{N}\right) + X_{d,\lambda_N}(j) \quad (j = 1, \ldots, N), \quad (5) \]

where \( m(x) \) is the unknown regression function and \( \{X_{d,\lambda_N}(j)\}_{j \in \mathbb{Z}} \) is a FLUR model. To estimate \( m(x) \), we consider the Priestley-Chao kernel estimator (Priestley and Chao 1972)

\[ \hat{m}(x) = \frac{1}{Nh} \sum_{j=1}^{N} K\left(\frac{Nj - j}{Nh}\right) Y(j). \quad (6) \]

In Theorems 2 and 3, we establish asymptotic normality, consistency, and the convergence rate for the Priestley-Chao kernel regression estimator under different values of \( \lambda_* \in [0, \infty] \). The results of Theorems 2 and 3 not only cover the asymptotic theory for nonparametric regression in Csörgő and Mielniczuk (1995a, b, c) when \( 0 < d < 1/2, \lambda_* = 0 \), but also extend to the case \( d > 0, \lambda_* \in (0, \infty) \). In the last case, the asymptotic results include a stochastic process which is called tempered fractional Brownian motion of the second kind (TFBMII).

In Sect. 4, we consider the parametric regression model

\[ Y(j) = \mu\left(\frac{j}{N}\right) + X_{d,\lambda_N}(j) \quad (j = 1, \ldots, N), \quad (7) \]

where \( \mu(x) \) is a continuous polynomial regression function with unknown knots and error process \( \{X_{d,\lambda_N}(j)\}_{j \in \mathbb{Z}} \). Theorems 5 and 6 establish the asymptotic distribution for the unknown parameters. Theorem 6 provides unified formulas for the asymptotic distribution of the least squares estimator of the unknown parameters based on tempered fractional calculus. We emphasize that, in our framework, we assume that the trend function \( \mu(x) \) is given by a piecewise polynomial and we aim to make an inference on the unknown parameters of the polynomial function. Therefore, we will make a distinction between our setting and the classical approximation or interpolation theory of splines or nonparametric spline smoothing (Wahba 1990; Boor 2001). In the latter references, an unknown regression function was estimated by splines. However, our work is more related with (non)linear regression and change point problems. For a general overview of nonlinear regression and change point problems, see Seber and Wild (1989) and Csörgő and Horváth (1997) respectively.

The paper is organized as follows. In Sect. 2, we briefly revise some important properties of TFBMII and its connection with tempered fractional calculus. In Sect. 3 we establish the finite dimensional distribution of normalized partial sums of weighted tempered linear processes and use it to develop the asymptotic theory for the nonparametric regression model in (5). In Sect. 4, we establish asymptotic results for the least squares estimator of the unknown knots of a piecewise polynomial regression model in (7). All proofs can be found in Appendix A.

In what follows, \( C \) denotes generic constants which may be different at different locations. We write \( \xrightarrow{\text{f.d.d.}} \) and \( \xrightarrow{\text{f.d.d.}} \) for convergence and equality of distributions in
the sense of finite-dimensional distributions respectively and \( \xrightarrow{d} \) to show the convergence in distribution. Denote \( \mathbb{N}_\pm := \{ \pm 1, \pm 2, \ldots \} \), \( \mathbb{R}_+ := (0, \infty) \), \( (x)_\pm := \max(\pm x, 0) \), \( x \in \mathbb{R} \), \( f := \int_{\mathbb{R}} \) and let \( L^p(\mathbb{R}) \) \((p \geq 1)\) denote the Banach space of measurable functions \( f : \mathbb{R} \to \mathbb{R} \) with finite norm \( \| f \|_p = \left( \int |f(x)|^p dx \right)^{1/p} \).

2 TFBMII: Definitions and main properties

TFBMII is introduced and discussed in detail in Sabzikar and Surgailis (2018a). For the sake of completeness, we give a short review of the definition and some essential properties of TFBMII that we will apply in the next sections.

Let \( d > -\frac{1}{2} \), \( \lambda > 0 \), and \( \{ B(t) \}_{t \in \mathbb{R}} \) be Brownian motion with mean zero and variance \( \sigma^2 |t| \). A TFBMII can be defined as the Wiener integral

\[
B^{II}_{d,\lambda}(t) := \frac{1}{\Gamma(d+1)} \int h_{d,\lambda}(t; y) dB(y),
\]

where the function \( y \mapsto h_{d,\lambda}(t; y) : \mathbb{R} \to \mathbb{R} \) is defined by

\[
h_{d,\lambda}(t; y) := (t-y)^d e^{-\lambda(t-y)} + (-y)^d e^{-\lambda(-y)} + \lambda \int_0^t (s-y)^d e^{-\lambda(s-y)} ds.
\]

Recall from Meerschaert and Sabzikar (2014) that the (positive and negative) tempered fractional integrals (TFI) and tempered fractional derivatives (TFD) of a function \( f : \mathbb{R} \to \mathbb{R} \) are defined by

\[
I_{\kappa,\lambda}^\pm f(y) := \frac{1}{\Gamma(\kappa)} \int f(s)(y-s)^{\kappa-1} e^{-\lambda(y-s)} ds, \quad \kappa > 0
\]

and

\[
D_{\kappa,\lambda}^\pm f(y) := \lambda^\kappa f(y) + \frac{\kappa}{\Gamma(1-\kappa)} \int (f(y) - f(s))(y-s)^{\kappa-1} e^{-\lambda(y-s)} ds
\]

for \( 0 < \kappa < 1 \) respectively. The TFI in (10) exists a.e. in \( \mathbb{R} \) for each \( f \in L^p(\mathbb{R}) \) and defines a bounded linear operator in \( L^p(\mathbb{R}) \), \( p \geq 1 \). The TFD in (11) exists for any absolutely continuous function \( f \in L^1(\mathbb{R}) \) such that \( f' \in L^1(\mathbb{R}) \); moreover, it can be extended to the fractional Sobolev space \( W^{\kappa,2}(\mathbb{R}) := \left\{ f \in L^2(\mathbb{R}) : \int (\lambda^2 + \omega^2)^\kappa |\hat{f}(\omega)|^2 d\omega < \infty \right\} \), see Meerschaert and Sabzikar (2014) for more properties of TFI and TFD. The function \( h_{d,\lambda}(t; y) \) can be written as

\[
\Gamma(d+1) I_{d,\lambda}^\pm 1_{[0,t]}(y) = h_{d,\lambda}(t; y), \quad d > 0,
\]

and

\[
\Gamma(d+1) D_{d,\lambda}^\pm 1_{[0,t]}(y) = h_{d,\lambda}(t; y), \quad -\frac{1}{2} < d < 0.
\]
Therefore, we can represent TFBMII in (8) as follows:

\[
B_{d,\lambda}^{II}(t) = \begin{cases} 
\int \mathbb{I}_{[0,t]}(y) dB(y), & d > 0, \\
\int \mathbb{I}_{[0,t]}(y) dB(y), & -\frac{1}{2} < d < 0.
\end{cases}
\]  

Using (14), Parseval’s formula for stochastic integrals (Samorodnitsky and Taqqu, 1994, Proposition 7.2.7) and the Fourier transform of TFI and TFD, we have the harmonizable representation of TFBMII

\[
B_{d,\lambda}^{II}(t) \overset{f.d.d.}{=} \frac{1}{\sqrt{2\pi}} \int e^{i\omega t} \frac{\lambda - 1}{i\omega} \frac{1}{(\lambda + i\omega)^{-d}} d\hat{B}(\omega),
\]

where \(\hat{B}\) is an even complex-valued Gaussian white noise, \(\hat{B}(dx) = \hat{B}(-dx)\) with zero mean and variance \(E|\hat{B}(dx)|^2 = dx\). The next proposition summarizes basic properties of \(B_{d,\lambda}^{II}(t)\). We refer the reader to Sabzikar and Surgailis (2018a) for more details.

**Proposition 1**

(i) TFBMII \(B_{d,\lambda}^{II}\) in (8) has stationary increments, such that

\[
\left\{ B_{d,\lambda}^{II}(ct) \right\}_{t \in \mathbb{R}} \overset{f.d.d.}{=} \left\{ c^{d+\frac{1}{2}} B_{d,c\lambda}^{II}(t) \right\}_{t \in \mathbb{R}},
\]

for any scale factor \(c > 0\) and is not a self-similar process.

(ii) TFBMII \(B_{d,\lambda}^{II}\) in (8) has a.s. continuous paths.

(iii) For \(d > 0\), the covariance function of TFBMII \(B_{d,\lambda}^{II}\) is given by

\[
E[B_{d,\lambda}^{II}(t)B_{d,\lambda}^{II}(s)] = C(d, \lambda) \int_0^t \int_0^s |u - v|^{d-\frac{1}{2}} K_{d-\frac{1}{2}}(\lambda|u - v|)dvdu,
\]

where \(C(d, \lambda) = \frac{2}{\sqrt{\pi} \Gamma(d+\lambda)^{d+\lambda}}\), \(d > 0\) and \(\lambda > 0\). Here \(K_{\nu}(x)\) is the modified Bessel function of the second kind (Abramowitz and Stegun, 1965, Chapter 9).

### 3 Nonparametric regression model in fixed equispaced design with FLUR errors

In this section, we study the asymptotic theory for the unknown regression function by using the Priestley-Chao kernel regression estimator when the errors has a FLUR or tempered linear model structure. More specifically, we consider the nonparametric equispaced fixed design regression model

\[
Y(j) = m\left(\frac{j}{N}\right) + X_{d,\lambda,N}(j) \quad (j = 1, \ldots, N),
\]
where \( \{ X_{d,\lambda_N}(j) \}_{j \in \mathbb{Z}} \) is satisfying (3)-(4). In addition, we assume that
\[
\sum_{k=0}^{\infty} k^j b_d(k) = 0, \quad 0 \leq j \leq [-d], \quad -\infty < d < 0,
\]
\[
\sum_{k=0}^{\infty} |b_d(k)| < \infty, \quad c_0 := \sum_{k=0}^{\infty} b_d(k) \neq 0, \quad d = 0
\]
(19)

Assumptions (19) and (20) are necessary for the validity of the convergence results.

The main statistic considered in our framework is
\[
S_{d,\lambda_N}(u) := \sum_{k=1}^{[Nu]} X_{d,\lambda_N}(k), \quad u \in [0, 1].
\]
(21)

We assume the innovations \( \{ \zeta(i) \}_{i \in \mathbb{Z}} \) in (3) are i.i.d with zero mean and unit variance so that \( N^{-1/2} \sum_{i=1}^{[Nt]} \zeta(i) \xrightarrow{d} B(t) \). Next, we state the required assumptions to obtain our main results in this section.

**Assumption 1** The tempering parameter \( \lambda \equiv \lambda_N \) may depend on \( N \) so that \( \lambda_N = o(1) \) and
\[
N\lambda \rightarrow \lambda_\ast \in [0, \infty]
\]
as \( N \rightarrow \infty \).

**Assumption 2** The bandwidth \( h = h_N \rightarrow 0 \) and \( Nh \rightarrow \infty \) as \( N \rightarrow \infty \).

**Assumption 3** Let \( K \) be a symmetric density function with support on \([-1, 1]\) with bounded first derivative \( K' \).

**Assumption 4** \( h \log(Nh) \rightarrow 0 \) as \( N \rightarrow \infty \)

**Theorem 1** Let \( \{ X_{d,\lambda_N}(j) \}_{j \in \mathbb{Z}} \) be tempered linear process satisfying (3)-(4) and (19)-(20). Then under Assumptions 1-3 and for \( 0 < x < 1 \), we have
(a) if \( K \in L^2(\mathbb{R}), \lambda_\ast = \infty, \) and \( d \in \mathbb{R} \setminus \mathbb{N}_- \), then
\[
\frac{\lambda N}{\sqrt{Nh}} \sum_{j=1}^{N} K\left( \frac{Nx - j}{Nh} \right) X_{d,\lambda_N}(j) \xrightarrow{f.d.d.} \sigma \int_0^2 K'(1-t) B(t) dt
\]
(23)
as \( N \rightarrow \infty \).

(b) if
\[
K \in \mathcal{A}_{d,0} := \left\{ f \in L^2(\mathbb{R}) : \int_{\mathbb{R}} |\hat{f}(\omega)|^2 |\omega|^{-2d} d\omega < \infty \right\},
\]
with \( \lambda_\ast = 0, -1/2 < d < 1/2 \), then
\[
\frac{1}{(Nh)^{d+1/2}} \sum_{j=1}^{N} K\left( \frac{Nx - j}{Nh} \right) X_{d,\lambda_N}(j) \xrightarrow{f.d.d.} \sigma \int_0^2 K'(1-t) B_{d,0}^H(t) dt
\]
(25)
as $N \to \infty$.

(c) If

$$K \in \mathcal{A}_{d, \lambda} := \{ f \in L^2(\mathbb{R}) : \int_\mathbb{R} (\lambda^2 + \omega^2)^{-d} |\hat{f}(\omega)|^2 \, d\omega < \infty \},$$

(26)

with $\lambda_* \in (0, \infty)$, and $d \in (0, \infty)$, then

$$\frac{1}{(Nh)^{d+1/2}} \sum_{j=1}^{N} K\left(\frac{Nx - j}{Nh}\right) X_{d, \lambda_N}(j) \xrightarrow{\text{f.d.d.}} \sigma \int_0^2 K'(1-t) B_{d, \lambda_*}^{II}(t) \, dt$$

(27)

as $N \to \infty$, where $B_{d, \lambda_*}^{II}$ is a TFBMII.

To estimate the regression function $m$, we consider the following kernel estimator Priestley and Chao (1972)

$$\hat{m}(x) = \frac{1}{Nh} \sum_{j=1}^{N} K\left(\frac{Nx - j}{Nh}\right) Y(j).$$

(28)

Next, we establish asymptotic normality, consistency, and convergence rate for the estimator (28), see Theorems 2 and 3.

**Theorem 2** Let $\{X_{d, \lambda_N}(j)\}_{j \in \mathbb{Z}}$ be tempered linear process satisfying (3)-(4) and (19)-(20). Assume the tempering parameter $\lambda$, the bandwidth $h$ and kernel $K$ satisfy Assumption 1-3. Further, let $K \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ and $\sigma^2$ is the variance of the innovations. Then, for fixed $x$, we have

(a) If $\lambda_* = \infty$ and $d \in \mathbb{R} \setminus \mathbb{N}_-$, then

$$\frac{\lambda_N^d}{\sqrt{Nh}} \sum_{j=1}^{N} K\left(\frac{Nx - j}{Nh}\right) X_{d, \lambda_N}(j) \xrightarrow{d} N(0, \sigma_{d, \infty}^2),$$

where

$$\sigma_{d, \infty}^2 = \sigma^2 \int_{-1}^{1} K^2(u) \, du$$

(29)

(b) If $-1/2 < d < 1/2$ and $\lambda_* = 0$, then

$$\frac{1}{(Nh)^{d+1/2}} \sum_{j=1}^{N} K\left(\frac{Nx - j}{Nh}\right) X_{d, \lambda_N}(j) \xrightarrow{d} N(0, \sigma_{d, 0}^2),$$

where $\sigma_{d, 0}^2 = \text{Cov}\left(\int_{\mathbb{R}} K(u) \, dB_{d, 0}(u), \int_{\mathbb{R}} K(v) \, dB_{d, 0}(v)\right)$. If we restrict $0 < d < 1/2$, then

$$\sigma_{d, 0}^2 = \sigma^2 \frac{c_0^2}{\Gamma^2(d)} \frac{2\Gamma\left(\frac{1}{2} - d\right)}{\sqrt{\pi} (d + \frac{1}{2}) 2^{2d+1} \Gamma(d+1)} \int_{-1}^{1} \int_{-1}^{1} K(u)K(v)|u-v|^{2d-1} \, du \, dv.$$  

(30)
(c) If \( d \in (0, \infty) \) and \( 0 < \lambda_* < \infty \), then
\[
\frac{1}{(Nh)^{d+1/2}} \sum_{j=1}^{N} K\left( \frac{N(x - j)}{Nh} \right) X_{d,\lambda,N}(j) \overset{d}{\to} N(0, \sigma_{d,\lambda_*}^2),
\]
where
\[
\sigma_{d,\lambda_*}^2 = \sigma^2 \frac{\Gamma(2d)}{\Gamma(d) \Gamma(2\lambda_*)^{d-1/2}} \int_{-1}^{1} K(u) K(v) |u - v|^{d-1/2} K_{d-1/2}(\lambda_* |u - v|) \, du \, dv
\]
where \( K_v(x) \) is the modified bessel function of the second kind and \( \sigma^2 \) is the variance of the innovations \( \{\xi \}_{j \in \mathbb{Z}} \).

**Theorem 3** Let the assumptions of Theorem 2 and Assumption 4 hold. For every \( k \in \mathbb{N} \) and \( 0 < x_1 < \ldots < x_k < 1 \), we have

(a) If \( \lambda_* = \infty \) and \( d \in \mathbb{R} \setminus \mathbb{N}_- \), then
\[
\lambda_N^d \sqrt{Nh} \left( \hat{m}(x_1) - \mathbb{E} \hat{m}(x_1), \ldots, \hat{m}(x_k) - \mathbb{E} \hat{m}(x_k) \right) \overset{d}{\to} \sqrt{\sigma_{d,\lambda_*}^2} \left( N_1, \ldots, N_k \right),
\]
where \( N_1, \ldots, N_k \) are independent standard normal distributions and \( \sigma_{d,\lambda_*}^2 \) is given by \((29)\).

(b) If \( -\frac{1}{2} < d < \frac{1}{2} \) and \( \lambda_* = 0 \), then
\[
(Nh)^{1/2-d} \left( \hat{m}(x_1) - \mathbb{E} \hat{m}(x_1), \ldots, \hat{m}(x_k) - \mathbb{E} \hat{m}(x_k) \right) \overset{d}{\to} \sqrt{\sigma_{d,0}^2} \left( N_1, \ldots, N_k \right),
\]
where \( N_1, \ldots, N_k \) are independent standard normal distributions and \( \sigma_{d,0}^2 \) is given by \((30)\).

(c) If \( d \in (0, \infty) \) and \( 0 < \lambda_* < \infty \), then
\[
(Nh)^{1/2-d} \left( \hat{m}(x_1) - \mathbb{E} \hat{m}(x_1), \ldots, \hat{m}(x_k) - \mathbb{E} \hat{m}(x_k) \right) \overset{d}{\to} \sqrt{\sigma_{d,\lambda_*}^2} \left( N_1, \ldots, N_k \right),
\]
where \( N_1, \ldots, N_k \) are independent standard normal distributions and \( \sigma_{d,\lambda_*}^2 \) is given by \((31)\).

**Remark 1** When \( \lambda_* = 0 \), the behaviour of \( S_{d,\lambda,N} \) is typical for long range dependency. This fact explains why part (b) of Theorem 3 is related with the asymptotic results for kernel regression function estimators obtained in Csörgő and Mielniczuk (1995b, Proposition 1) and Deo (1997, Theorem 2) with long range dependence errors.

**Remark 2** In case of long range dependence and twice continuously differentiable regression function, Hall and Hart (1990) and Beran and Feng (2013) state that the minimax optimal convergence rate for the regression estimator \( \hat{m}(x) \) is of order \( n^{-r} \),
with \( r = 2(1 - d)/(5 - 2d) \), uniformly for \( d \in (-1/2, 1/2) \) under certain regularity conditions. Ray and Tsay (1997, Eq. 4) showed that the optimal bandwidth \( h_{\text{opt}} \) for long range dependence (\( d \in (0, 1/2) \)) is

\[
h_{\text{opt}} = \left[ \frac{(1 - 2d)\sigma^2 \Gamma(1 - 2d) \int_{-1}^{1} \int_{-1}^{1} K(u)K(v) |u - v|^{2d-1} \, du \, dv}{\Gamma(d) \Gamma(1 - d) \left[ \int_{-1}^{1} u^2 K(u) du \right]^2} \left[ \int_{0}^{1} \left( \int_{0}^{1} K(u) \, du \right)^2 \, dx \right] \right]^{1/5} n^{-2d/5}.
\]

In the tempered case, the tempering parameter \( \lambda \) effects the optimal bandwidth and rate of convergence of the estimator. It would be interesting to consider the optimal bandwidth and the convergence rate when the tempering parameter \( \lambda \) is sample size dependent (i.e. \( \lambda_N \)). For the latter, we conjecture that the bandwidth is \( O(n^{-1/5}) \), which is of the same order as short range dependence and i.i.d. case. However, a rigorous theoretical and simulation study would be needed and is beyond the scope of this paper. Also, to the authors’ knowledge, a method to estimate \( \lambda_N \) does not exist. A possible lead could be found in De Brabanter et al. (2018). By using a kernel function \( K \) such that \( K(0) = 0 \), an asymptotically optimal bandwidth can be obtained by minimizing the residual sums of squares without any prior knowledge of the error process. We believe that this method could also be used to directly estimate the tempering parameter \( \lambda_N \) from data.

### 4 Piecewise polynomial regression model with semi-long memory error

In this section we investigate the asymptotic theory for piecewise polynomial and spline regression with partially unknown knots and errors having a FLUR or tempered linear model. We will obtain unified formulas for the asymptotic distribution of least squares estimators of the unknown parameters based on tempered fractional calculus. More precisely, consider the parametric regression model

\[
Y(j) = \mu\left(\frac{j}{N}\right) + X_{d,\lambda_N}(j) \quad (j = 1, \ldots, N),
\]

where \( \mu(\cdot) \) is a continuous polynomial function such that

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

and \( \{X_{d,\lambda}(j)\} \) is a tempered linear processes satisfying (3)-(4) and (19)-(20). Here \( a = (a_1, \ldots, a_p) \) denotes unknown regression coefficients and \( f_1, \ldots, f_p \) are truncated power spline basis functions defined as \( f_1(s) = 1, f_2(s) = s, \ldots, f_q(s) = s^{q-1}, f_{q+1}(s) = (s-\eta)_+, \ldots, f_p(s) = (s-\eta)^{p-q}_+ \) with \( (s-\eta)_+ := 1_{\{s-\eta > 0\}} \) and \( \eta \) is an unknown knot parameter. The regression function \( \mu(s) = \sum_{i=1}^{p} a_i f_i(s) = \mu(s; \theta) \) depends on a \((p+1)\)-dimensional parameter vector \( \theta = (a^T, \eta)^T \in \Theta = \mathbb{R}^p \times (0, 1). \)
Since we want the model to be identifiable, we assume in addition \( a_i \neq 0 \) for at least one \( i \in \{q + 1, \ldots, p \} \).

Let \( \theta \) be the ordinary least squares estimator of \( \theta \) that minimizes

\[
\sum_{j=1}^{N} \left[ Y(j) - \mu \left( \frac{j}{N}, \theta \right) \right]^2
\]

with respect to \( \theta \in \Theta \). Let \( \eta \in (0, 1) \) and define the \( n \times p \) matrix

\[
W_N = W_N(\eta) = (w_{j,i})_{j=1,\ldots,N; i=1,\ldots,p} = (w_{N,1}, \ldots, w_{N,p})
\]

with \( w_{j,i} = f_i(\frac{j}{N}) (1 \leq j \leq N; 1 \leq i \leq p) \), and column vectors denoted by \( w_{N,i}(i = 1, \ldots, p) \). According to Beran et al. (2014), \( W_N^T W_N \) is invertible for large \( N \) and the projection matrix onto the column space of \( W_N(\eta) \) is

\[
P_{W_N} = P_{W_N(\eta)} = W_N(W_N^T W_N)^{-1} W_N^T.
\]

Hence, for given observations \( X = (X_1, \ldots, X_n)^T \), \( \hat{\eta} \) is obtained by minimizing \( \beta = \|X - P_{W_N(\hat{\eta})}X\|_2 \) with respect to \( \hat{\eta} \). Now, one can compute \( \hat{\theta} = (\hat{a}_1, \ldots, \hat{a}_p) \) by projecting \( X \) onto the column space of the design matrix \( W_N(\hat{\eta}) \). For the estimated mean function we have

\[
\left[ \mu \left( \frac{1}{N}, \hat{\theta} \right), \mu \left( \frac{2}{N}, \hat{\theta} \right), \ldots, \mu \left( 1, \hat{\theta} \right) \right]^T = P_{W_N(\hat{\eta})}X = P_{W_N(\hat{\eta})}[\mu_N(\theta) + e_N],
\]

where \( \mu_N(\theta) = [\mu(h, \theta)]_{j=1,\ldots,N} \) and \( e_N = (X_{d,\lambda_1}(1), \ldots, X_{d,\lambda_N}(N))^T \). In general, the partial derivatives \( \partial_t \mu(t; \eta, a) \) and \( \partial_\eta \mu(t; \eta, a) \) do not exist if \( \eta = t \). However, one can use the left and right derivatives of \( \mu(t; \eta, a) \) since they exist everywhere. We denote the left and right partial derivatives of \( \mu \) with respect to \( \theta_t \) by \( \mu_{(i-)} \) and \( \mu_{(i+)} \) respectively. Partial derivatives in the sense of an absolutely continuous function will be denoted by \( \mu_{(i)} \). Note that \( \mu_{(i+)} = \mu_{(i-)} = \mu_{(i)} \) almost everywhere.

Defining the \( N \times (p + 1) \) matrix

\[
M_{N+} = [\mu_{(i+)}(t/N)]_{t=1,\ldots,n; i=1,\ldots,p+1} \in \mathbb{R}^{N \times (p+1)},
\]

we have

\[
\lim_{N \to \infty} N^{-1}(M_{N+}^T M_{N+})_{jk} = \int_0^1 \mu_{(j)}(s, \theta) \mu_{(k)}(s, \theta) \, ds.
\]

Similarly to \( W_N^T W_N \), the matrix \( M_{N+}^T M_{N+} \) has full rank for large \( N \) such that

\[
A = \lim_{N \to \infty} N (M_{N+}^T M_{N+})^{-1}
\]

is well defined. Theorem 4 shows that \( \lambda_N^{d} \sqrt{N} (\hat{\theta} - \theta) \) and \( N^{\frac{1}{2} - d} (\hat{\theta} - \theta) \) are asymptotically equivalent for different values of \( \lambda_N \) to \( \lambda_N^{d} \sqrt{N} (M_{N+}^T M_{N+})^{-1} M_{N+}^T e_N \) and
\[ N^{1/2-d} (M_{N+}^T M_{N+})^{-1} M_{N+}^T e_N \] respectively. Theorems 5 and 6 establish the asymptotic distribution of the parameter vector \( \hat{\theta} \) for different values of \( \lambda_* \).

**Theorem 4** Let \( \{X_d, \lambda_N(t)\}_{t \in \mathbb{Z}} \) be a tempered linear process given by (3). Then for any \( \Delta > 0 \)

(a) If \( \lambda_* = \infty \), \( d > 0 \), and \( \lambda_N = o(N^{-1/(2-2d)}) \), then

\[
\mathbb{P}\left( \lambda_N^d \sqrt{N} \left\| \hat{\theta} - \theta - (M_{N+}^T M_{N+})^{-1} M_{N+}^T e_N \right\| > \Delta \right) = o(1), \tag{35}
\]

as \( N \to \infty \),

(b) If \( \lambda_* \in [0, \infty) \) and \( 0 < d < 1/2 \), then

\[
\mathbb{P}\left( N^{1/2-d} \left\| \hat{\theta} - \theta - (M_{N+}^T M_{N+})^{-1} M_{N+}^T e_N \right\| > \Delta \right) = o(1), \tag{36}
\]

as \( N \to \infty \), where \( \| \cdot \| \) denotes the Euclidean norm.

**Theorem 5** Let \( \{X_d, \lambda_N(t)\}_{t \in \mathbb{Z}} \) be a tempered linear process given by (3).

(a) If \( d \in \mathbb{R} \setminus \mathbb{N}_- \) and \( \lambda_* = \infty \), then

\[
\lambda_N^d \sqrt{N} (\hat{\theta} - \theta) \xrightarrow{f.d.d.} \Lambda \Xi,
\]

where \( \Lambda \) is given by (34), \( \Xi = \left[ \int_{\mathbb{R}} \mu_{(i+)}(s) dB(s) \right]_{i=1, \ldots, p+1} \) is a random vector process, and \( B(s) \) is a Brownian motion.

(b) If \( 0 < d < 1/2 \) and \( \lambda_* = 0 \), then

\[
N^{1/2-d} (\hat{\theta} - \theta) \xrightarrow{f.d.d.} \Lambda \Xi,
\]

where \( \Lambda \) is given by (34), \( \Xi = \left[ \int_{\mathbb{R}} \mu_{(i+)}(s) dB_{d,0}^H(s) \right]_{i=1, \ldots, p+1} \) is a random vector process, and \( B_{d,0}^H \) is a multiple of FBM.

(c) If \( 0 < d < 1/2 \) and \( \lambda_* \in (0, \infty) \), then

\[
N^{1/2-d} (\hat{\theta} - \theta) \xrightarrow{f.d.d.} \Lambda \Xi,
\]

where \( \Lambda \) is given by (34), \( \Xi = \left[ \int_{\mathbb{R}} \mu_{(i+)}(s) dB_{d,\lambda_*}^H(s) \right]_{i=1, \ldots, p+1} \) is a random vector process, and \( B_{d,\lambda_*}^H \) is TFBMII.

The following theorem shows that the limit distribution in Theorem 5 is Gaussian and also gives the closed form of the covariance matrix of the random vector \( \Xi \).

**Theorem 6** Under the assumptions of Theorem 5, the limit distribution of the random vector \( \Lambda \Xi \) is Gaussian with zero mean and covariance \( \Lambda \Sigma_0 \Lambda \). That is

\[
N^{1/2-d} (\hat{\theta} - \theta) \xrightarrow{f.d.d.} \Lambda \Xi \sim N(0, \Lambda \Sigma_0 \Lambda)
\]
as \( N \to \infty \). Moreover,

(a) for \( d \in \mathbb{R} \setminus \mathbb{N}_- \) and \( \lambda_* = \infty \), we have

\[
\Sigma_\infty = \sigma^2 \left[ \int_{\mathbb{R}} \int_{\mathbb{R}} \mu_{(i+)}(t) \mu_{(k+)}(s) ds \ dt \right]_{i,k=1,\ldots,p+1}. \tag{37}
\]

(b) for \( 0 < d < 1/2 \) and \( \lambda_* = 0 \), we have

\[
\Sigma_0 = \sigma^2 \left[ \int_{\mathbb{R}} \left( \Re_{d,0} \mu_{(i+)} \right)(s) \left( \Re_{d,0} \mu_{(k+)} \right)(s) \ ds \right]_{i,k=1,\ldots,p+1}. \tag{38}
\]

Moreover the covariance (38) can be written as

\[
\Sigma_0 = \sigma^2 \frac{c_0^2}{\Gamma^2(d)} \frac{2 \Gamma\left(\frac{1}{2} - d\right)}{\sqrt{\pi} (d + \frac{1}{2}) 2^{2d+1} \Gamma(d + 1)} \int_{\mathbb{R}} \mu_{(i+)}(t) \mu_{(k+)}(s) |t - s|^{2d-1} ds \ dt. \tag{39}
\]

(c) for \( d > 0 \) and \( \lambda_* \in (0, \infty) \), we have

\[
\Sigma_{\lambda_*} = \frac{c_0^2}{\Gamma^2(d)} \left[ \int_{\mathbb{R}} \left( \Re_{d,\lambda_*} \mu_{(i+)} \right)(s) \left( \Re_{d,\lambda_*} \mu_{(k+)} \right)(s) \ ds \right]_{i,k=1,\ldots,p+1}. \tag{40}
\]

Moreover the covariance (40) can be written as

\[
\Sigma_{\lambda_*} = \sigma^2 \frac{c_0^2}{\Gamma^2(d)} \frac{2}{\Gamma(d) \sqrt{\pi} (2\lambda^*)^{d-\frac{1}{2}}} \int_{\mathbb{R}} \mu_{(i+)}(t) \mu_{(k+)}(s) |t - s|^{d-\frac{1}{2}} K_{d-\frac{1}{2}}(\lambda_*|t - s|) ds \ dt. \tag{41}
\]

**Remark 3**

(a) Parts (a) and (b) of Theorem 6 are related to the result in Beran et al. (2014, Theorem 3.3) for the short and long memory cases.

(b) Theorem 6 is also valid for \(-\frac{1}{2} < d < 0\) and \( \lambda_* \in [0, \infty) \). In this case,

\[
\Sigma_0 = \sigma^2 \left[ \int_{\mathbb{R}} \left( \Re_{-d,\lambda_*} \mu_{(i+)} \right)(s) \left( \Re_{-d,\lambda_*} \mu_{(k+)} \right)(s) \ ds \right]_{i,k=1,\ldots,p+1}. \tag{42}
\]

However, a closed form for \( \Sigma_0 \) for the case \( d > 0 \) does not exist.

**Remark 4**

All results of this section consider one unknown knot. However, similar results can be obtained for an arbitrary continuous piecewise polynomial function

\[
\mu(s) = \sum_{k=0}^l \sum_{j=1}^{p_k} a_{k,j} (s - \eta_{k})^{b_{j,k}}
\]
(a) [Kernel density estimate of $\frac{\lambda_d}{\sigma_{d,\infty} \sqrt{Nh}} \sum_{j=1}^{N} K(\frac{N x - j}{Nh}) X_d,\lambda_N (j)$ (full) and standard normal density (dashed). a $d = 0.1$, $\lambda_N = N^{-0.3}$, $h = 0.0691$; b $d = 0.7$, $\lambda_N = N^{-0.3}$, $h = 0.1172$; with $b_{j,k} < b_{j+1,k}$, knots $0 = \eta_0 < \eta_1 < \cdots < \eta_l < 1$ of which some (but not necessarily all) are unknown, and the condition $b_{j,k} \geq 1$ for $k \geq 1$ (needed for continuity).

5 Simulations

In this section we investigate the finite sample performance of asymptotic results in Theorem 2. We consider three different experiments all based on $N = 500$ and 20,000 replications. In all of the experiments we consider the regression model $Y_j = m(\frac{j}{N}) + X_d,\lambda_N (j)$, $j = 1, \ldots, N$ with $m(\frac{j}{N}) = 10 \sin(\pi \frac{j}{N})$. The bandwidth for all three experiments is determined by minimizing

$$\frac{1}{N} \sum_{j=1}^{N} \left[ \hat{m}(\frac{j}{N}) - m(\frac{j}{N}) \right]^2$$

where $\hat{m}(-)$, given by (6) with Epanechnikov kernel, is estimated at $x = 0.5$. We use the R package artfima (McLeod et al. 2016) to simulate the tempered linear process (3). The innovations have mean zero and unit variance. Figure 2 illustrates the finite sample approximation for Theorem 2(a) for two different values of the parameter $d$. The values of $c_0$ are estimated from the Monte Carlo simulation. Figure 3 and 4 illustrate the finite sample approximation for Theorem 2(b) and 2(c) respectively.

6 Conclusion

Using the idea of tempered fractional calculus coming from mathematical modeling in probability, we construct a fractional local to unity root (FLUR) model which is an extension of the well known local to unity root model in economics. Next, we used
the FLUR model to develop the asymptotic theory in parametric and nonparametric regression models in statistics. Our theory provides a broad inclusive approach to regression theory and accommodates both short and long memory innovations as well as intermediate mechanisms in which long memory properties may become attenuated at very long lags according to the presence of a tempering parameter. Allowing the tempering parameter to be sample size dependent opens up a further range of potential time series behavior. One advantage of this extension is that under certain conditions the limit theory results in a pivotal statistic, thereby simplifying inference. This feature leads to opportunities for further research on procedures that can free empirical investigators from having to estimate memory parameters, a property that is likely to be extremely useful in multivariate cases.

**Declarations**

**Conflict of interest** The authors declare that they have no conflict of interest.
A Proofs

Before we prove the main results of the paper, we first state two technical lemmas upon which our results are based. Lemma 1 and Lemma 2, play an important role in establishing the asymptotic results in Sect. 4. Next, we introduce some notations that will be used in Lemma 2.

For the function \( f \) and \( m \in \mathbb{N} \cup \{ \infty \} \), we define the approximation

\[
 f_{N,m}^+(y) = \sum_{j=0}^{m} f\left(\frac{j}{N}\right) 1_{\left]\frac{j}{N}, \frac{j+1}{N}\right]}(y), \quad f_{N,m}^- = \sum_{j=-m}^{-1} f\left(\frac{j}{N}\right) 1_{\left]\frac{j}{N}, \frac{j+1}{N}\right]}(y),
\]

\[
 f_N^+ = f_{N,\infty}^+, \quad f_N^- = f_{N,\infty}^-, \quad f_N = f_N^+ + f_N^-.
\]

Lemma 1 Let \( b_d(k) \) be defined as in (4). Then for any \( y \in \mathbb{R} \)

\[
 \left(\lambda_N + i \frac{y}{N}\right)^d \sum_{k=0}^{\infty} e^{-(\lambda_N + i \frac{y}{N})k} \ b_d(k) \sim 1
\]

as \( N \to \infty \).

**proof Lemma 1** For \( d > 0 \), since \( \sum_{k=0}^{N} b_d(k) \sim 1/(d \Gamma(d))N^d \) as \( N \to \infty \) according to (4) and \( e^{-(\lambda_N + \frac{ik}{N})} \leq 1 \), then according to the Tauberian theorem for power series (Feller 1971, p. 447 Theorem 5) we have

\[
 \sum_{k=0}^{\infty} e^{-(\lambda_N + \frac{ik}{N})k} \ b_d(k) \sim (1 - e^{-(\lambda_N + \frac{ik}{N})k})^{-d} \text{ as } N \to \infty
\]

and consequently \( \left(\lambda_N + i \frac{y}{N}\right)^d/(1 - e^{-(\lambda_N + i \frac{y}{N})k})^d \sim 1 \) for \( \lambda_N \to 0 \) and \( N \to \infty \) as \( N \to \infty \), proving (43). For \( -1 < d < 0 \), define \( \tilde{b}_d(k) = \sum_{i=k}^{\infty} b_d(i) \sim -1/(d \Gamma(d))k^{d-1} \) with \( \tilde{d} = d + 1 \in (0, 1) \). Next, we have that

\[
 \tilde{b}_d(0) = \sum_{i=0}^{\infty} b_d(i) = \sum_{i=0}^{k-1} b_d(i) + \sum_{i=k}^{\infty} b_d(i) = 0
\]

and therefore \( \sum_{i=0}^{k-1} b_d(i) = -\sum_{i=k}^{\infty} b_d(i) \). Using summation by parts (Giraitis et al., 2012, p. 32, Eq. 2.5.8) yields

\[
 \lim_{s \to \infty} \sum_{j=0}^{s} e^{-(\lambda_N + i \frac{y}{N})j} \ b_d(j) = \lim_{s \to \infty} \left[ e^{-(\lambda_N + i \frac{y}{N})s} \sum_{j=0}^{s} b_d(j) \right.
\]

\[
 \left. + \sum_{j=0}^{s-1} e^{-(\lambda_N + i \frac{y}{N})s} - e^{-(\lambda_N + i \frac{y}{N})(s+1)} \sum_{i=0}^{j} b_d(i) \right]
\]
= 0 + \{1 - e^{-\left(\lambda_N + \frac{j}{N}\right)}\} \lim_{s \to \infty} \sum_{j=0}^{s-1} e^{-\left(\lambda_N + \frac{j}{N}\right)} j \sum_{i=0}^{j} b_d(i).

By setting \( j = k - 1 \) and using (44) we have

\[
\begin{align*}
\lim_{s \to \infty} \sum_{j=0}^{s} e^{-\left(\lambda_N + \frac{j}{N}\right)} j b_d(j) &= \{1 - e^{-\left(\lambda_N,\lambda\right)}\} \lim_{s \to \infty} \sum_{k=1}^{s-1} e^{-\left(\lambda_N + \frac{k}{N}\right)} \sum_{i=k}^{k-1} b_d(i) \\
&= -e^{\left(\lambda_N + \frac{j}{N}\right)} \{1 - e^{-\left(\lambda_N + \frac{j}{N}\right)}\} \lim_{s \to \infty} \sum_{k=1}^{s-1} e^{-\left(\lambda_N + \frac{k}{N}\right)} \sum_{i=k}^{\infty} b_d(i) \\
&= -e^{\left(\lambda_N + \frac{j}{N}\right)} \{1 - e^{-\left(\lambda_N + \frac{j}{N}\right)}\} \lim_{s \to \infty} \sum_{k=1}^{s-1} e^{-\left(\lambda_N + \frac{k}{N}\right)} \sum_{i=k}^{\infty} b_d(i).
\end{align*}
\]

Application of the Tauberian theorem for power series (Feller, 1971, p. 447 Theorem 5) yields

\[
\sum_{k=0}^{\infty} e^{-\left(\lambda_N + \frac{k}{N}\right)} b_d(k) \sim \{1 - e^{-\left(\lambda_N + \frac{j}{N}\right)}\}^{1-d} = \{1 - e^{-\left(\lambda_N + \frac{j}{N}\right)}\}^{d}
\]
as \( N \to \infty \), proving (43). In the general case \(-j < d < -j + 1, j = 1, 2, \ldots \) (43) follows similarly using summation by parts \( j \) times. For \( d = 0 \), it can be shown that the same result holds under an additional assumption on the sum of the \( b_d(k) \)'s Sabzikar and Surgailis (2018b).

**Lemma 2** Let \( \{X_{d,\lambda_N}(j)\}_{j \in \mathbb{Z}} \) be tempered linear process given by (3), \( N\lambda_N \to \lambda_\ast \in (0, \infty) \) and \( d > -1/2 \). Let \( \mathcal{A}_{d,\lambda_\ast} \) be the class of functions defined by (26) and let

**Condition A**: \( f, f_{\pm} \in \mathcal{A}_{d,\lambda_N}, \|f_{\pm} - f_{N,m}\|_{\mathcal{A}_{d,\lambda_N}} \to 0 \), as \( m \to \infty \), \( \|f - f_N\|_{\mathcal{A}_{d,\lambda_N}} \to 0 \), as \( N \to \infty \)

be satisfied, then

\[
\frac{1}{N^{d+1/2}} \sum_{j=-\infty}^{\infty} f\left(\frac{j}{N}\right) X_{d,\lambda_N}(j) \xrightarrow{f.d.d.} \int_{\mathbb{R}} f(u) \, dB_{d,\lambda_\ast}^H(u)
\]
as \( N \to \infty \).

**Proof Lemma 2** We first note that \( \{X_{d,\lambda_N}(j)\}_{j \in \mathbb{Z}} \) can be written as

\[
X_{d,\lambda_N}(j) = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} e^{i\omega j} \sum_{k=0}^{\infty} e^{-i\omega k} e^{-\lambda_N k} b_d(k) \hat{B}(d\omega),
\]

\( \hat{B} \) Springer
where $\hat{B}(d\omega)$ is complex-valued Gaussian noise with $\mathbb{E}|\hat{B}(d\omega)|^2 = d\omega$, see Brockwell and Davis (2012, Sects. 4.6–4.7). Define

$$U_N = \frac{1}{N^{d+\frac{1}{2}}} \sum_{j=-\infty}^{\infty} f\left(\frac{j}{N}\right) X_{d,\lambda_N}(j), \quad U = \int_{\mathbb{R}} f(u) \, dB_{d,\lambda_\ast}^I(u). \quad (47)$$

The Wiener integral $U$ is well-defined, since $f \in \mathcal{A}_{d,\lambda}$. To show that the series $U_N$ is well-defined in the $L^2(\Omega)$, first apply the spectral representation of \{\{\{X_{d,\lambda_N}(j)\}\}j\in\mathbb{Z}$ given by (46)

$$\frac{1}{N^{d+\frac{1}{2}}} \sum_{j=0}^{m} f\left(\frac{j}{N}\right) X_{d,\lambda_N}(j) = \frac{1}{N^{d+\frac{1}{2}}} \int_{-\pi}^{\pi} \left[ \sum_{j=0}^{m} \frac{1}{\sqrt{2\pi}} f\left(\frac{j}{N}\right) e^{ij\omega} \right] \sum_{k=0}^{\infty} e^{-(\lambda_N+i\omega)k} b_d(k) \, d\hat{B}(\omega)$$

$$= \frac{1}{N^{d+\frac{1}{2}}} \int_{\mathbb{R}} \left[ \sum_{j=0}^{m} \frac{1}{\sqrt{2\pi}} f\left(\frac{j}{N}\right) e^{ij\frac{\omega}{N}} \right] 1_{[-N\pi,N\pi]}(y)$$

$$\times \sum_{k=0}^{\infty} e^{-(\lambda_N+i\frac{\omega}{N})k} b_d(k) d\hat{B}(N^{-1}y) \quad (48)$$

$$= \frac{1}{N^{d+\frac{1}{2}}} \int_{\mathbb{R}} \left[ \sum_{j=0}^{m} \frac{1}{\sqrt{2\pi}} f\left(\frac{j}{N}\right) e^{ij\frac{(j+1)\omega}{N}} - e^{ij\frac{\omega}{N}} \right]$$

$$\times \frac{i\frac{\omega}{N}}{e^{\frac{\omega}{N}} - 1} 1_{[-N\pi,N\pi]}(y) \sum_{k=0}^{\infty} e^{-(\lambda_N+i\frac{\omega}{N})k} b_d(k) d\hat{B}(N^{-1}y)$$

$$= \frac{1}{N^{d+\frac{1}{2}}} \int_{\mathbb{R}} \hat{f}_{N,m}(y) \frac{i\frac{\omega}{N}}{e^{\frac{\omega}{N}} - 1} \sum_{k=0}^{\infty} e^{-(\lambda_N+i\frac{\omega}{N})k} b_d(k) d\hat{B}(N^{-1}y),$$

where $\hat{f}_{N,m}(y) = \sum_{j=0}^{m} f\left(\frac{j}{N}\right) \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{i\omega y} 1_{\left(\frac{j}{N}, \frac{j+1}{N}\right)}(\omega) \, d\omega$ is the Fourier transform of $f_{N,m}$. We note

$$\sum_{k=0}^{\infty} e^{-(\lambda_N+i\omega)k} b_d(k) < C \left(\lambda_N + \frac{i\omega}{N}\right)^{-d}, \quad (49)$$

for $d > -\frac{1}{2}$ and a constant $C$ by Lemma 1. Using (48) and (49), we have

$$\mathbb{E} \left| \frac{1}{N^{d+\frac{1}{2}}} \sum_{j=0}^{m} f\left(\frac{j}{N}\right) X_{d,\lambda_N}(j) \right|^2$$

$$= \int_{\mathbb{R}} \left| \hat{f}_{N,m}(y) \right|^2 \left| \frac{i\frac{\omega}{N}}{e^{\frac{\omega}{N}} - 1} \right|^2 \frac{1}{N^{2d}} \sum_{k=0}^{\infty} e^{-(\lambda_N+i\frac{\omega}{N})k} b_d(k) \right|^2 \, dy$$

$\square$ Springer
\[
\leq \frac{\pi^2}{4} C \int_{\mathbb{R}} |\hat{f}_{N,m}(y)|^2 \left[ (N\lambda_N)^2 + y^2 \right]^{-d} dy
= C' \|f_{N,m}\|^2_{A_{3,\lambda_N}},
\]

(50)

where \(C'\) is another constant. Now, for \(m_2 > m_1 \geq 1\), we have

\[
\mathbb{E} \left\| \frac{1}{N^{d+\frac{1}{2}}} \sum_{j=m_1+1}^{m_2} f\left(\frac{j}{N}\right) X_{d,\lambda_N}(j) \right\|^2 \leq C' \|f_{N,m_2}^+ - f_{N,m_1}^+\|^2_{A_{3,\lambda_N}} \to 0
\]
as \(m_1, m_2 \to \infty\) and this shows the series is well-defined. The following remark illustrates the inequality in (50).

**Remark 5** In (50) we used Lemma 1 and \(\left\| \frac{iy}{e^{i\pi} - 1} \right\|^2 \leq \frac{\pi^2}{4}\) for \(y \in [-N\pi, N\pi]\). This can be seen as follows

\[
\left| \frac{iy}{e^{i\pi} - 1} \right|^2 = \frac{y^2}{N^2} \frac{\cos \frac{y}{N} + i \sin \frac{y}{N} - 1}{2(1 - \cos \frac{y}{N})} = \frac{1}{4} \frac{\sin^2 \frac{y}{2N}}{\sin^2 \frac{y}{2N}}.
\]

Then taking the limit yields

\[
\lim_{N \to \infty} \frac{1}{4} \frac{N^2 \pi^2}{\sin^2 \frac{\pi N}{2N}} = \frac{1}{4} \lim_{N \to \infty} \frac{\pi^2}{\sin^2 \frac{\pi}{2}} = \frac{\pi^2}{4}.
\]

Next, we show that \(U_N\) converges in distribution to \(U\) as \(N \to \infty\). Similar to Meerschaert and Sabzikar (2014, Theorem 3.15), the set of elementary functions are dense in \(A_{d,\lambda}\) and then there exists a sequence of elementary functions \(f^l\) such that \(\|f - f^l\|_{A_{3}} \to 0\), as \(l \to \infty\). Now, assume

\[
U_N^l = \frac{1}{N^{d+\frac{1}{2}}} \sum_{j=-\infty}^{\infty} f^l\left(\frac{j}{N}\right) X_{d,\lambda_N}(j), \quad U^l = \Gamma^{-1}(d + 1) \int_{\mathbb{R}} f^l(u) dB_{d,\lambda_N}^U(u).
\]

(51)

Observe that \(U_N^l\) is well-defined, since \(U_N^l\) has a finite number of terms and the elementary function \(f^l\) is in \(A_{3}\). According to Meerschaert and Sabzikar (1968, Theorem 4.2.), the series \(U_N\) converges in distribution to \(U\) if

Step 1 \(U^l \xrightarrow{d} U\), as \(l \to \infty\),

Step 2 for all \(l \in \mathbb{N}\), \(U_N^l \xrightarrow{d} U^l\), as \(N \to \infty\),

Step 3 \(\lim_{N \to \infty} \lim_{l \to \infty} \mathbb{E} \left| U_N^l - U_N \right|^2 = 0\).

Step 1: The random variables \(U^l\) and \(U\) have normal distribution with mean zero and variances \(\|f^l\|_{A_{3,\lambda_N}}\) and \(\|f\|_{A_{3,\lambda_N}}\), respectively, since \(f\) and \(f^l\) are in \(A_{3,\lambda_N}\). Therefore \(\mathbb{E} \left| U^l - U \right|^2 = \|f^l - f\|_{A_{3,\lambda_N}} \to 0\) as \(l \to \infty\).
Step 2: Note that $f^l$ is an elementary function and hence $U_N^l$, given by (51), can be written as $U_N^l = \frac{1}{N^{d+\frac{1}{2}}} \int_{\mathbb{R}} f^l(u) dS_{d, \lambda_N}(u)$. Now, apply part (iii) of Theorem 4.3 in Sabzikar and Surgailis (2018b) to see that $\frac{S_{d, \lambda_N}(u)}{N^{d+\frac{1}{2}}} \xrightarrow{\text{f.d.d.}} \Gamma^{-1}(d+1)B_{d, \lambda_*}^H(u)$, as $N \to \infty$, and this implies that $U_N^l \xrightarrow{\text{f.d.d.}} U^l$, as $N \to \infty$.

Step 3: By a similar arguments of (49) and (50), we have

$$
\mathbb{E} \left| U_N^l - U_N^l \right|^2 = \int_{\mathbb{R}} \left| \hat{f}_N^l(y) - \hat{f}_N^l(y) \right|^2 \left| \frac{iy}{N} \right|^2 \left| 1 - e^{-(\lambda_N + \frac{i}{N})} \right|^{-2d} dy 
\leq C \int_{\mathbb{R}} \left| \hat{f}_N^l(y) - \hat{f}_N^l(y) \right|^2 \left[ (N\lambda_N)^2 + y^2 \right]^{-d} dy,
$$

(52)

where $\hat{f}_N^l(y)$ and $\hat{f}_N^l(y)$ are the Fourier transforms of

$$
f_N^l(u) := \sum_{j=0}^{\infty} f^l \left( \frac{j}{N} \right) 1 \left( \frac{j}{N}, \frac{N}{N} - j \right) (u)
$$

and $f_N := \sum_{j=0}^{\infty} f \left( \frac{j}{N} \right) 1 \left( \frac{j}{N}, \frac{N}{N} - j \right) (u)$ respectively. Note that $f^l$ is an elementary function and therefore $\hat{f}_N^l$ converges to $\hat{f}^l$ at every point and $\left| \hat{f}_N^l(\omega) - \hat{f}_N^l(\omega) \right| \leq \hat{g}^l(\omega)$ uniformly in $N$, for some function $\hat{g}^l(\omega)$ which is bounded by the minimum of $C_1$ and $C_2|\omega|^{-1}$ for all $\omega \in \mathbb{R}$ (see Theorem 3.2. in Pipiras and Taqqu (2000) for more details). Let $\mu_{d, \lambda}(d\omega) = (\lambda^2 + \omega^2)^{-d} d\omega$ be the measure on the real line for $d > -\frac{1}{2}$, then $\hat{g}^l(\omega) \in L^2(\mathbb{R}, \mu_{d, \lambda})$. Now apply the Dominated Convergence Theorem to see that

$$
\| f_N^l - f_N^l \|_{A_3}^2 = \| \hat{f}_N^l - \hat{f}_N^l \|_{L^2(\mathbb{R}, \mu_{d, \lambda})}^2 \to 0,
$$

(53)
as $N \to \infty$. From (48) and (53), we have

$$
\mathbb{E} \left| U_N^l - U_N^l \right|^2 \leq C \| \hat{f}_N^l - \hat{f}_N^l \|_{A_3}^2 
\leq C \left[ \| \hat{f}_N^l - \hat{f}_N^l \|_{A_3}^2 + \| \hat{f} - \hat{f}_N^l \|_{A_3}^2 + \| \hat{f} - \hat{f}_N^l \|_{A_3}^2 \right].
$$

The first two terms tend to zero as $N \to \infty$ because of (53) and Condition A respectively, and the last term tends to zero as $l \to 0$ (see step 1) and this completes the proof of Step 3. \hfill \Box

**proof Theorem 1** We prove only part (c) and omit the proofs of parts (a) and (b) due to the similarity of proofs. We first show that

$$
\frac{1}{(Nh)^{d+\frac{1}{2}}} \sum_{j=1}^{N} K \left( \frac{Nj - j}{Nh} \right) X_{d, \lambda_N}(j) \xrightarrow{\text{f.d.d.}} \frac{1}{\Gamma(d+1)} \int_{0}^{2} K'(1-t)B_{d, \lambda_*}^H(t)dt,
$$

(54)
where $B_{d,\lambda_*}^I(t)$ is TFBMII. Starting from the l.h.s of (54), using Riemann sums to integrals and integration by parts

\[
\frac{1}{(Nh)^{d+\frac{1}{2}}} \sum_{j=1}^{N} K\left(\frac{N(x-j)}{Nh}\right) X_{d,\lambda_N}(j) = \frac{1}{(Nh)^{d+\frac{1}{2}}} \int_0^1 K\left(\frac{x-y}{h}\right) d S_{d,\lambda_N}(y)
\]

\[
= \frac{1}{h(Nh)^{d+\frac{1}{2}}} \int_0^1 K'\left(\frac{x-y}{h}\right) S_{d,\lambda_N}(y) dy
\]

\[
= \frac{1}{(Nh)^{d+\frac{1}{2}}} \int_{-1}^{1} K'(u) S_{d,\lambda_N}(x-hu) du
\]

\[
= \frac{1}{(Nh)^{d+\frac{1}{2}}} \int_{-1}^{1} K'(u) \sum_{j=\lfloor N(x-hu) \rfloor}^{\lfloor N(x-hu) \rfloor - 1} X_{d,\lambda_N}(j) du,
\]

where we used

\[
S_{d,\lambda_N}(x-hu) = \sum_{j=1}^{\lfloor N(x-hu) \rfloor - 1} X_{d,\lambda_N}(j) + \sum_{j=\lfloor N(x-hu) \rfloor}^{\lfloor N(x-hu) \rfloor - 1} X_{d,\lambda_N}(j)
\]

and the assumptions on the kernel function $K$ to see that

\[
\sum_{j=1}^{\lfloor N(x-hu) \rfloor - 1} X_{d,\lambda_N}(j) = \int_{-1}^{1} K'(u) du = 0.
\]

Next, by stationarity of $X_{d,\lambda_N}(j)$ and a change of variable we have

\[
\sum_{j=\lfloor N(x-hu) \rfloor}^{\lfloor N(x-hu) \rfloor - 1} X_{d,\lambda_N}(j) = \sum_{j=1}^{\lfloor N(x-hu) \rfloor - \lfloor N(x-h) \rfloor + 1} X_{d,\lambda_N}(j)
\]

\[
\sum_{j=\lfloor N(x-h) \rfloor}^{\lfloor N(x-hu) \rfloor - \lfloor N(x-h) \rfloor + 1} X_{d,\lambda_N}(j) = \sum_{j=1}^{\lfloor N(x-hu) \rfloor - \lfloor N(x-h) \rfloor + 1} X_{d,\lambda_N}(j),
\]

where $l_x(u) = \lfloor N(x-hu) \rfloor - \lfloor N(x-h) \rfloor + 1$. Consequently, from (55) and (56), we get

\[
\frac{1}{(Nh)^{d+\frac{1}{2}}} \int_{-1}^{1} K'(u) \sum_{j=\lfloor N(x-h) \rfloor}^{\lfloor N(x-hu) \rfloor} X_{d,\lambda_N}(j) du
\]

\[
\stackrel{f.d.d.}{=} \frac{1}{(Nh)^{d+\frac{1}{2}}} \int_{-1}^{1} K'(u) \sum_{j=1}^{l_x(u)} X_{d,\lambda_N}(j) du
\]
According to Sabzikar and Surgailis (2018b, Theorem 4.3) we have

\[
\frac{1}{(Nh)^{d+\frac{1}{2}}} \sum_{j=1}^{\lfloor Nh \rfloor} X_{d,\lambda_N}(j) \int_0^{\lfloor Nh \rfloor} K'(1-t)X_{d,\lambda_N}(j)dt + o_p(1)
\]

(57)

in \(D[0, 2]\) provided \(Nh \to \infty\). Now, the desired result (54) follows from (57), (58), and the continuous mapping theorem.

**proof of Theorem 2** The proof of this theorem follows by Proposition 1 and Theorem 1 and hence we omit the details.

**proof of Theorem 3** For brevity, we restrict the proof of the theorem to \(k = 2\). Moreover, we just prove part (c) of the theorem since the proofs of the other two cases are similar. We first note that for each \(0 < x_i < 1\),

\[
\hat{A}_{N,i} = (Nh)^{1-d}\left[\hat{m}(x_i) - \mathbb{E}\hat{m}(x_i)\right] = \frac{1}{(Nh)^{d+\frac{1}{2}}} \sum_{1 \leq s \leq N} K\left(\frac{N(x_i - s)}{Nh}\right)X_{d,\lambda_N}(s).
\]

(59)

Let \(j_i\) be integers such that \(|N(x_i - j_i)| \leq 1\) for \(i = 1, 2\) and (similar to Deo (1997)) define

\[
\tilde{A}_{N,i} = \frac{1}{(Nh)^{d+\frac{1}{2}}} \sum_{s = \lfloor j_i \rfloor - \lfloor Nh \rfloor}^{\lfloor j_i \rfloor + \lfloor Nh \rfloor} K\left(\frac{j_i - s}{Nh}\right)X_{d,\lambda_N}(s)
\]

(60)

for \(i = 1, 2\). Since the kernel function \(K\) vanishes in \(\mathbb{R} \setminus [-1, 1]\) and \(|K'(x)| \leq C\) for all \(x \in [-1, 1]\), it follows that

\[
\hat{A}_{N,i} - \tilde{A}_{N,i} = o_p(1)
\]

(61)

for \(i = 1, 2\). By a change of variable \(s = \nu + j_1 - Nh\) and \(s = \nu + j_2 - Nh - \lfloor N\delta \rfloor\), with \(\delta = x_2 - x_1\), for \(\hat{A}_{N,1}\) and \(\tilde{A}_{N,2}\) respectively, use the fact that \(X_{d,\lambda_N}\) is stationary, and \(|K'(x)| \leq C\) to see that

\[
\hat{A}_{N,1}, \hat{A}_{N,2} \overset{f.d.d.}{=} \left(\hat{A}_{N,1}^*, \hat{A}_{N,2}^*\right) + o_p(1),
\]

(62)

where

\[
\hat{A}_{N,1}^* = \sum_{v=1}^{2\lfloor Nh \rfloor} K\left(\frac{v}{Nh} - 1\right)X_{d,\lambda_N}(v)
\]

(63)

and

\[
\hat{A}_{N,2}^* = \sum_{v=[N\delta]}^{\lfloor Nh \rfloor} K\left(\frac{v - \lfloor N\delta \rfloor}{Nh} - 1\right)X_{d,\lambda_N}(v).
\]

(64)
We use the partial sums $A_{N,i}$, for $i = 1, 2$, to establish the functional limit theorems. Let $\{K_m\}$ be a sequence of elementary functions such that $K_m \to K$ in $L^2$ as $m \to \infty$. Define $A_{m,N,i}$ be as (63) and (64) with $K_m(x) = \sum_{i=1}^{m} a_i \mathbf{1}(t_{i-1}, t_i)(x)$, where $a_i$ are some constants and $-1 \leq t_i \leq 1$ for $i = 0, \ldots, m$. We can rewrite $A_{m,N,i}$ as

$$A_{m,N,i} = 2^{d+\frac{1}{2}} \int_0^2 K_m(u - 1)dS_{N,i}^*(u) + o_P(1), \tag{65}$$

where

$$S_{N,1}^*(s) = \frac{1}{(2Nh)^{d+\frac{1}{2}}} \sum_{t=1}^{[Nh]s} X_{d,\lambda_N}(t) \tag{66}$$

and

$$S_{N,2}^*(s) = \frac{1}{(2Nh)^{d+\frac{1}{2}}} \sum_{t=1}^{[Nh]s} X_{d,\lambda_N}(t + [N\delta]). \tag{67}$$

Using Sabzikar and Surgailis (2018b, Theorem 4.3) and the continuous mapping theorem yields

$$A_{m,N,i} \overset{f.d.}{\to} A_m = \int_0^2 K_m(u - 1)dB_{d,\lambda_N}^H(u), \tag{68}$$

as $N \to \infty$ and hence

$$\sigma_{ii}^2 = \int_0^2 \int_0^2 K_m(u - 1)K_m(v - 1)\text{Cov}\left(B_{d,\lambda_N}^H(u), B_{d,\lambda_N}^H(v)\right)du \, dv \tag{69}$$

Next, we need to show that $A_{m,N,1}$ and $A_{m,N,2}$ are asymptotically independent (i.e. $\sigma_{12}^2 = \sigma_{21}^2 = 0$). Observe that

$$A_{m,N,1} = 2^{d+\frac{1}{2}} \sum_{j=1}^{m} a_j [S_{N1}^*(t_j) - S_{N1}^*(t_{j-1})] + o_P(1) \tag{70}$$

and

$$S_{N,1}^*(t_j) - S_{N,1}^*(t_{j-1}) = \frac{1}{(2Nh)^{d+\frac{1}{2}}} \sum_{s=[Nh]t_{j-1}}^{[Nh]t_j} X_{d,\lambda_N}(s) = \sum_{p=-\infty}^{\infty} d_{pN} \xi(p), \tag{71}$$

where

$$d_{pN} = \frac{1}{(2Nh)^{d+\frac{1}{2}}} \sum_{t=[Nh]t_{j-1}}^{[Nh]t_j} b_d(p - t)e^{-\lambda_N(p-t)}. \tag{72}$$
Since \( b_d(j) e^{-\lambda_N j} \sim C j^{d-1} e^{-\lambda_N j} \) for large \( j \), see (4), then for \( p > \lfloor \lceil Nh \rceil t_j \rfloor \), we have
\[
|d_{pN}| \leq C |Nh|^{d+\frac{1}{2}} Nh(p - \lfloor \lceil Nh \rceil t_j \rfloor)^{d-1} e^{-\lambda_N(p - \lfloor \lceil Nh \rceil t_j \rfloor)},
\]
where \( C \) is a constant. Therefore, we get
\[
\lim_{N \to \infty} \sum_{|p| > M} d_{pN}^2 = 0,
\]
since \( h \log(Nh) \to 0 \) and \( M = Nh \log(Nh) \). Consequently,
\[
\lim_{N \to \infty} \sum_{|p| \leq M} d_{pN} \xi(p) = 0
\]
and by a similar argument
\[
\lim_{N \to \infty} \sum_{|p| \leq M} d_{pN} \xi(p + \lfloor N\delta \rfloor) = 0
\]
From (75), (76), and \( \lfloor N\delta \rfloor - 2N \to \infty \), we conclude that \( S_{N,1}^*(t_j) - S_{N,1}^*(t_{j-1}) \) and \( S_{N,2}^*(t_{j'}) - S_{N,2}^*(t_{j'-1}) \) are asymptotically independent for all \( j, j' \) and this implies that \( A_{m,N,1}^* \) and \( A_{m,N,2}^* \) are asymptotically independent. Thus
\[
\left( A_{m,N,1}^*, A_{m,N,2}^* \right) \xrightarrow{f.d.d.} N^2(0, \Sigma),
\]
where \( \sigma_{ii}^2 \) is given by (69) and \( \sigma_{12}^2 = \sigma_{21}^2 = 0 \). Since
\[
\lim_{m \to \infty} \sigma_{ii}^2 = \int_0^2 \int_0^2 K(u - 1)K(v - 1)|u - v|^{d-\frac{1}{2}} K_{d-\frac{1}{2}}(\lambda_* |u - v|) du dv
\]
and by Pipiras and Taqqu (1997, Theorem 2), one obtains
\[
\lim_{m \to \infty} \lim_{N \to \infty} \text{Var}(A_{m,N,1}^* - A_{N,1}^*) = 0, \quad \text{for } i = 1, 2,
\]
then the desired results follows by (61), (62), (78), and (79).

**Proof of Theorem 4**: The proofs of part (a) and part (b) are similar and we just prove part (b) for \( \lambda_* \in (0, \infty) \). Using the triangle inequality yields
\[
\mathbb{P}\left( N^{\frac{1}{2}-d} \| \hat{\theta} - \theta - (M_{N+}^T M_{N+})^{-1} M_{N+}^T e_N \| > \Delta \right)
\leq \mathbb{P}\left( N^{\frac{1}{2}-d} \| \hat{\theta} - \theta \| > \frac{\Delta}{2} \right) + \mathbb{P}\left( N^{\frac{1}{2}-d} \| (M_{N+}^T M_{N+})^{-1} M_{N+}^T e_N \| > \frac{\Delta}{2} \right).
\]
For the first term, we have by Weiershäuser (2012, p. 95 Theorem 5.1.9), (3) and Markov’s inequality,

$$\mathbb{P}\left(N^{\frac{1}{2} - d} \left\| \hat{\theta} - \theta \right\| > \frac{\Delta}{2}\right) \leq \frac{4 \mathbb{E} \left\| \hat{\theta} - \theta \right\|^2}{\Delta^2 N^{1 - 2d}} = O\left[\min\left(N^{-\beta} N^{2d - 1}, N^{A_d - 2}\right)\right].$$

Since $\Delta$ is arbitrary, $\lim_{N \to \infty} \mathbb{P}\left(N^{\frac{1}{2} - d} \left\| \hat{\theta} - \theta \right\| > \frac{4}{2}\right) = 0$ for $d < 1/2$. For the second term and again by Markov’s inequality

$$\mathbb{P}\left(N^{\frac{1}{2} - d} \left\| (M_{N+}^T + M_{N+}^T)^{-1} M_{N+}^T e_N \right\| > \frac{\Delta}{2}\right) \leq \frac{4 \mathbb{E} \left\| (M_{N+}^T + M_{N+}^T)^{-1} M_{N+}^T e_N \right\|^2}{\Delta^2 N^{1 - 2d}}.$$

(80)

Let $\Omega = (M_{N+}^T + M_{N+}^T)^{-1} M_{N+}^T$, then $\mathbb{E} \left\| \Omega e_N \right\|^2 = tr(\Omega \Sigma e_N e_N) + \{\mathbb{E}(e_N)^T \Sigma e_N e_N \mathbb{E}(e_N)^T \Sigma e_N e_N \}$. The proofs of part (a) and part (b) are similar and hence we just give the proof for part (b). We first note that $\int_\mathbb{R} \mathbb{E}(e_N)^T \Sigma e_N e_N \mathbb{E}(e_N)^T \Sigma e_N e_N \, du$, $\lambda(\cdot)$, are tempered mean zero linear processes. Since $(X_{d, \lambda}(t))_{t \in \mathbb{Z}}$ is a tempered mean zero linear process we have $\mathbb{E} \left\| \Omega e_N \right\|^2 = \mathbb{E}(\Omega \Sigma e_N e_N)$. Further, the variance-covariance matrix of $e_N$ is finite, see (17). Consequently, the numerator of (80) is finite since $M_{N+}^T + M_{N+}^T$ is full rank for $N \to \infty$ and hence the second term goes to zero for $d < 1/2$. \hfill \Box

**proof of Theorem 5** The proofs of part (a) and part (b) are similar and hence we just give the proof for part (b). We first note that $\int_\mathbb{R} \mathbb{E}(e_N)^T \Sigma e_N e_N \, du$, $\lambda(\cdot)$, are tempered mean zero linear processes. Since $(X_{d, \lambda}(t))_{t \in \mathbb{Z}}$ is a tempered mean zero linear process we have $\mathbb{E} \left\| \Omega e_N \right\|^2 = \mathbb{E}(\Omega \Sigma e_N e_N)$. Further, the variance-covariance matrix of $e_N$ is finite, see (17). Consequently, the numerator of (80) is finite since $M_{N+}^T + M_{N+}^T$ is full rank for $N \to \infty$ and hence the second term goes to zero for $d < 1/2$. \hfill \Box

We have

$$N^{\frac{1}{2} - d} (M_{N+}^T + M_{N+}^T)^{-1} M_{N+}^T e_N \xrightarrow{\text{f.d.d.}} \Lambda \left[ \int_\mathbb{R} \mathbb{E}(\mathbb{E}(e_N)^T \Sigma e_N e_N \mathbb{E}(e_N)^T \Sigma e_N e_N) \right]_{i=1, \ldots, p+1}$$

(81)

as $N \to \infty$. Observe that $N(M_{N+}^T + M_{N+}^T)^{-1} \Lambda \to \Lambda$ as $N \to \infty$. Therefore the RHS of (81) follows if we show

$$\frac{1}{N^{d + \frac{1}{2}}} M_{N+}^T e_N \xrightarrow{\text{f.d.d.}} \left[ \int_\mathbb{R} \mathbb{E}(\mathbb{E}(e_N)^T \Sigma e_N e_N \mathbb{E}(e_N)^T \Sigma e_N e_N) \right]_{i=1, \ldots, p+1}.$$

But this is equivalent to show that

$$\frac{1}{N^{d + \frac{1}{2}}} \langle \alpha, M_{N+}^T e_N \rangle \to \left[ \int_\mathbb{R} \mathbb{E}(\mathbb{E}(e_N)^T \Sigma e_N e_N \mathbb{E}(e_N)^T \Sigma e_N e_N) \right]_{i=1, \ldots, p+1}, \alpha \in \mathbb{R}^{p+1}.\ \ (82)$$

Note that

$$\langle \alpha, M_{N+}^T e_N \rangle = \sum_{i=1}^{p+1} \sum_{j=1}^{n} \alpha_i \mu(i+)(\frac{j}{N}) X_{d, \lambda_N}(j) \ \ (83)$$
and by Lemma 2

\[ N^{-(d+1)/2} \sum_{i=1}^{p+1} \sum_{j=1}^{N} \alpha_i \mu_{i(j)} \left( \frac{j}{N} \right) X_{d, \lambda, k}(j) \xrightarrow{\text{f.d.d.}} \int_{\mathbb{R}} m_\alpha(u) d B_{d, \lambda, k}^I(u) \]

where \( m_\alpha(u) := \sum_{i=1}^{p+1} \alpha_i \mu_{i(u)}(u) \) and this completes the proof. \( \square \)

**proof of Theorem 6** We only proof part (c) since the other parts follow a similar procedure. Let \( \Xi \) be the random vector

\[ \Xi = \left[ \int_{\mathbb{R}} \mu_{i(u)}(s) d B_{d, \lambda, k}^I(s) \right]_{i=1, \ldots, p+1}. \]  

(84)

Then we can write

\[ \int_{\mathbb{R}} \mu_{i(u)}(s) d B_{d, \lambda, k}^I(s) = \int_{\mathbb{R}} \left( \left[ \begin{array}{c} 1 \end{array} \right] \mu_{i(u)}(s) \right) d B(s) \]

(85)

for \( i = 1, \ldots, p+1 \). We observe that \( \int_{\mathbb{R}} \left( \left[ \begin{array}{c} 1 \end{array} \right] \mu_{i(u)}(s) \right) d B(s) \) is a Gaussian stochastic process with mean zero and finite variance \( \int_{\mathbb{R}} \left( \left[ \begin{array}{c} 1 \end{array} \right] \mu_{i(u)}(s) \right)^2 d s \). Using the Itô-isometry for the Wiener integrals, one can see \( \Xi \) has the covariance matrix

\[ \Sigma_0 = \left[ \int_{\mathbb{R}} \left( \left[ \begin{array}{c} 1 \end{array} \right] \mu_{i(u)}(s) \right) \left( \left[ \begin{array}{c} 1 \end{array} \right] \mu_{k(u)}(s) \right) d s \right]_{i, k=1, \ldots, p+1} \]

(86)

and consequently \( \Lambda \Xi \) has normal distribution with covariance matrix \( \Lambda \Sigma_0 \Lambda \) and this completes the proof of the first part. Next, we have

\[
\int_{\mathbb{R}} \left( \left[ \begin{array}{c} 1 \end{array} \right] \mu_{i(u)}(s) \right) \left( \left[ \begin{array}{c} 1 \end{array} \right] \mu_{k(u)}(s) \right) d s = \int_{\mathbb{R}} \mathcal{F} \left[ \left( \left[ \begin{array}{c} 1 \end{array} \right] \mu_{i(u)}(s) \right) \omega \right] \mathcal{F} \left[ \left( \left[ \begin{array}{c} 1 \end{array} \right] \mu_{k(u)}(s) \right) \omega \right] d \omega \\
= \int_{\mathbb{R}} \int_{\mathbb{R}} \bar{\mu}_{i(u)}(\omega) \bar{\mu}_{k(u)}(\omega) (\lambda_*^2 + \omega^2)^{-d} d \omega \\
= \int_{\mathbb{R}} \mu_{i(u)}(t) \mu_{k(u)}(s) \int_{\mathbb{R}} e^{i \omega (t-s)} (\lambda_*^2 + \omega^2)^{-d} d \omega ds dt \\
= 2 \int_{\mathbb{R}} \mu_{i(u)}(t) \mu_{k(u)}(s) \int_{0}^{\infty} \cos(\omega (t-s)) (\lambda_*^2 + \omega^2)^{-d} d \omega ds dt \\
= C \int_{\mathbb{R}} \mu_{i(u)}(t) \mu_{k(u)}(s) |t-s|^{-d/2} K_{d/2-1} (\lambda_* |t-s|) ds dt,
\]

(87)

where \( C = \frac{2}{\Gamma(d) \sqrt{\pi} (2\lambda)^{d-1/2}} \) and we used

\[
\int_{0}^{\infty} \frac{\cos(\omega x)}{(\lambda^2 + x^2)^{v+1/2}} dx = \frac{\sqrt{\pi}}{\Gamma(v + 1/2)} \left( \frac{|x|}{2\lambda} \right)^v K_v(\lambda |x|)
\]

(88)
for $\nu > \frac{1}{2}$ and $\lambda > 0$ in (87). This completes the proof of the second part and Theorem. "\hfill \square"

References

Abramowitz M, Stegun I (1965) Handbook of mathematical functions, 9th edn. Dover, Illinois
Baillie R (1996) Long-memory processes and fractional integration in econometrics. J Econ 73:5–59
Barnsden-Nielsen O (1998) Processes of normal inverse gaussian type. Finance Stoch 2:41–68
Beran J, Feng Y (2013) Optimal convergence rates in non-parametric regression with fractional time series errors. J Time Ser Anal 34:30–39
Beran J, Weiershäuser A, Galizia C, Rein J, Smith B, Strauch M (2014) On piecewise polynomial regression under general dependence conditions, with an application to calcium-imaging data. Sankhya B 76:49–81
Billingsley P (1968) Convergence of probability measures. Wiley, Hoboken
Bobkoski M (2011) Hypothesis testing in nonstationary time series. Dissertation, University of Wisconsin
Boor C (2001) A practical guide to splines. Springer, Berlin
Brockwell P, Davis R (2012) Time series: theory and methods, 2nd edn. Springer, Berlin
Chan N, Wei C (1987) Asymptotic inference for nearly nonstationary AR(1) processes. Ann Statist 15:1050–1063
Csörgő M, Horváth L (1997) Limit theorems in change-point analysis. Wiley, Hoboken
Csörgő S, Mielniczuk J (1995) Close short-range dependent sums and regression estimation. Acta Sci Math (Szeged) 60:177–196
Csörgő S, Mielniczuk J (1995) Nonparametric regression under long-range dependent normal errors. Ann Statist 23:1000–1014
Csörgő S, Mielniczuk J (1995) Distant long-range dependent sums and regression estimation. Stochastic Process Appl 59:143–155
Dacorogna M, Moller U, Nagler R, Olsen R, Pictet O (1993) A geographical model for the daily and weekly seasonal volatility in the foreign exchange market. J Int Money Finance 12:413–443
De Brabanter K, Cao F, Gijbels I, Opsomer J (2018) Local polynomial regression with correlated errors in random design and unknown correlation structure. Biometrika 105:681–690
Deo R (1997) Nonparametric regression with long-memory errors. Stat Probab Lett 33:89–94
Feller W (1971) An introduction to probability theory and its applications, 2nd edn. Wiley, Hoboken
Giraitis L, Kokoszka P, Leipur R (2000) Stationary arch models: dependence structure and central limit theorem. Econ Theory 16:3–22
Giraitis L, Kokoszka P, Leipur R, Teyssiére G (2003) On the power of R/S-type tests under contiguous and semi-long memory alternatives. Acta Appl Math 78:285–299
Giraitis L, Koul H, Surgailis D (2012) Large sample inference for long memory processes. Imperial College Press, London
Granger C, Joyeux R (1980) An introduction to long-memory time series models and fractional differencing. J Time Ser Anal 1:15–29
Granger C, Ding Z (1996) Varieties of long memory models. J Econ 73:61–77
Guo H, Koul H (2007) Nonparametric regression with heteroscedastic long memory errors. J Statist Plann Inference 137:379–404
Hall P, Hart J (1990) Nonparametric regression with long-range dependence. Stochastic Process Appl 36:339–351
Hosking J (1981) Fractional differencing. Biometrika 68:165–176
Hosking J (1984) Modeling persistence in hydrological time series using fractional differencing. Water Resour Res 20:1898–1908
McLeod AI, Meerschaert MM, Sabzikar F (2016) Arttima v1.5. https://cran.r-project.org/web/packages/arttima/arttima.pdf
Meerschaert M, Sabzikar F (2014) Stochastic integration with respect to tempered fractional brownian motion. Stochastic Process Appl 124:2363–2387
Müller U, Watson M (2014) Measuring uncertainty about long-run predictions. Rev Econ Stud 83:1711–1740
Phillips P (1987) Time series regression with a unit root. Econometrica 55:277–301
Phillips P, Yu J (2011) Dating the timeline of financial bubbles during the subprime crisis. Quantitat Econ 2:455–491
Phillips P, Shi S, Yu J (2015) Testing for multiple bubbles: historical episodes of exuberance and collapse in the S&P 500. Int Econ Rev 56:1042–1076
Phillips P, Shi S, Yu J (2015) Testing for multiple bubbles: limit theory of real time detectors. Int Econ Rev 56:1077–1131
Pipiras V, Taqqu M (1997) Asymptotic theory for certain regression models with long memory errors. J Time Ser Anal 18:385–393
Pipiras V, Taqqu M (2000) Convergence of weighted sums of random variables with long range dependence. Stochast Process Appl 90:157–174
Priestley M, Chao M (1972) Nonparametric function fitting. J R Stat Soc Ser B Stat Methodol 34:385–392
Ray B, Tsay R (1997) Bandwidth selection for kernel regression with long-range dependent errors. Biometrika 84:791–802
Robinson P (1997) Large sample inference for nonparametric regression with dependent errors. Ann Stat 25:2054–2083
Sabzikar F, Surgailis D (2018) Tempered fractional brownian and stable motions of second kind. Stat Probab Lett 132:17–27
Sabzikar F, Surgailis D (2018) Invariance principles for tempered fractionally integrated processes. Stochast Process Appl 128:3419–3438
Samorodnitsky G, Taqqu M (1994) Stable non-gaussian random processes: stochastic models with infinite variance. Chapman and Hall, Boca Raton
Seber G, Wild C (1989) Nonlinear regression. Wiley, Hoboken
Wahba G (1990) Spline models for observational data. Society for Industrial and Applied Mathematics (SIAM), New Delhi
Weiershäuser A (2012) Piecewise polynomial regression with fractional residuals for the analysis of calcium imaging data. Dissertation. http://kops.uni-konstanz.de/handle/123456789/18867, University of Konstanz

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.