The Annals of Statistics
2016, Vol. 44, No. 5, 2058–2088
DOI: 10.1214/15-AOS1402

LOCAL INTRINSIC STATIONARITY AND ITS INFERENCE

BY TAILEN HSING*, THOMAS BROWN† AND BRIAN THELEN*

University of Michigan and Exponent*†

Dense spatial data are commonplace nowadays, and they provide the impetus for addressing nonstationarity in a general way. This paper extends the notion of intrinsic random function by allowing the stationary component of the covariance to vary with spatial location. A nonparametric estimation procedure based on gridded data is introduced for the case where the covariance function is regularly varying at any location. An asymptotic theory is developed for the procedure on a fixed domain by letting the grid size tend to zero.

1. Local intrinsic stationarity. The spatial process of interest in this paper is denoted by $Y(\mathbf{t})$ where $\mathbf{t} \in$ some $\Omega \subset \mathbb{R}^d$ for an arbitrary dimension $d = 1, 2, \ldots$. Let

$$\mu(\mathbf{t}) = \mathbb{E}Y(\mathbf{t}) \quad \text{and} \quad C(\mathbf{t}, \mathbf{s}) = \operatorname{Cov}\bigl(Y(\mathbf{t}), Y(\mathbf{s})\bigr).$$

Common goals of spatial statistics include the inference of μ and C and the prediction of $Y(\mathbf{t})$ based on observations $Y(\mathbf{t}_1), \ldots, Y(\mathbf{t}_n)$. In doing so, Y is often assumed to be second-order stationary, namely, $\mu(\mathbf{t})$ is constant and $C(\mathbf{s}, \mathbf{t})$ is a function of $\mathbf{t} - \mathbf{s}$. However, (global) stationarity is a restrictive assumption which is sometimes hard to justify. In this paper, we focus on the modeling and inference of covariance nonstationarity. There has been considerable progress recently on this topic; see a review in Sampson (2010). Not surprisingly, nonstationary models are often created with (latent) stationary processes as building blocks. For completeness, we briefly mention a few such examples here. Higdon, Swall and Kern (1999), Fuentes (2001) and Fuentes (2002) consider nonstationary processes that are convolutions of deterministic kernels with stationary processes. Sampson and Guttorp (1992), Anderes and Stein (2008) and Anderes and Chatterjee (2009) introduce nonstationarity by composing a stationary process with a spatial deformation transformation. The nonstationary process in Fuglstad et al. (2013) and Fuglstad et al. (2015) is the solution of an SPDE driven by a Gaussian white noise;

Tribute: Professor Peter Hall had a huge influence on our research. The approach of combining nonparametric smoothing and spatial statistics in this paper was directly motivated by some of his results. His body of work that spans almost the entire field of statistics will be a profound and lasting source of inspiration for statistical research. We will also miss Professor Hall dearly for his extraordinary leadership and the very personal ways in which he helped so many of us.

Received October 2014; revised September 2015.

MSC2010 subject classifications. 60G12, 62M30, 62G05, 62G20.

Key words and phrases. Intrinsic random functions, nonparametric estimation, nonstationary spatial process.

a result of Whittle (1954) provides the motivation for that approach. Finally, Kim, Mallick and Holmes (2005) models nonstationarity by joining stationary processes defined on sub-regions of Ω, where the choice of sub-regions as well as the inference of the process are conducted using a Bayesian approach.

The purpose of this paper is to develop a general framework for nonstationarity when dense, gridded spatial data are available. Data of this type are common nowadays. Some climate and geospatial data are gridded data on the sphere but those with high resolution can be reasonably treated as being on the plane over small regions; see data available at NASA's Earth Science or the U.S. Geological Survey websites. Compared with existing approaches, represented by those mentioned in the previous paragraph, the approach that we introduce in this paper aims at directly modeling the covariance locally and does not involve globally defined latent processes. Our model is very general and is satisfied by many nonstationary models in the literature. The features of a statistical model, as a rule, are closely aligned with the proposed inference approach. The inference approach of this paper is nonparametric smoothing. In that regard, the development below bears some resemblance to Anderes and Stein (2008) and Anderes and Chatterjee (2009).

We start by reviewing a well-known covariance model for spatial processes. For $\boldsymbol{\ell} = (\ell_1, \dots, \ell_d) \in \{0, 1, 2, \dots\}^d$, and $\mathbf{h} = (h_1, \dots, h_d) \in \mathbb{R}^d$, let $|\boldsymbol{\ell}| = \sum_{j=1}^d \ell_j$ and $\mathbf{h}^{\boldsymbol{\ell}} = h_1^{\ell_1} \cdots h_d^{\ell_d}$. There should be no confusion that $|\cdot|$ will also denote the usual absolute value as well as the norm for $\mathbb{R}^d$. Write

$$C(\mathbf{t}, \mathbf{s}) = \sum_{|\boldsymbol{\ell}|=0}^{r} b_{\boldsymbol{\ell}}(\mathbf{t})\mathbf{s}^{\boldsymbol{\ell}} + \sum_{|\boldsymbol{\ell}|=0}^{r} b_{\boldsymbol{\ell}}(\mathbf{s})\mathbf{t}^{\boldsymbol{\ell}} + R(\mathbf{t}, \mathbf{s}) \tag{1}$$

for some r, where the $b_{\boldsymbol{\ell}}$ are arbitrary measurable functions and the summations are over all $\boldsymbol{\ell}$ with $|\boldsymbol{\ell}| = 0, \dots, r$. A process $Y(\mathbf{t})$ is said to be an intrinsic random function of order r [Matheron (1964, 1973)], denoted as IRF-r, if

$$R(\mathbf{s}, \mathbf{t}) = K(\mathbf{t} - \mathbf{s}) \tag{2}$$

for some function K, referred to as generalized covariance. The most familiar intrinsic random functions are the IRF-0's, commonly known as intrinsically stationary processes, for which stationarity can be achieved by one-step differencing and the generalized covariance K is the negative semivariogram. An obvious example of an intrinsically stationary process is the standard Brownian motion. If Y is an IRF-r and $(\lambda_i, \mathbf{t}_i)$, $1 \le i \le m$, are such that

$$\sum_{i=1}^{m} \lambda_i \mathbf{t}_i^{\boldsymbol{\ell}} = 0, \qquad |\boldsymbol{\ell}| = 0, \dots, r,$$

then

$$\sum_{i=1}^{m}\sum_{j=1}^{m} \lambda_i \lambda_j C(\mathbf{t}_i, \mathbf{t}_j) = \sum_{i=1}^{m}\sum_{j=1}^{m} \lambda_i \lambda_j K(\mathbf{t}_i - \mathbf{t}_j).$$

Thus, K can assume the role of C in kriging when focusing on predictors of the form $\sum_{i=1}^{m} \lambda_i Y(\mathbf{t}_i)$. For more details on intrinsic random functions and the corresponding kriging problem, see Cressie (1993), Stein (1999) and Chilès and Delfiner (2012).

While the notion of intrinsic random function extends stationarity in meaningful ways, it is still quite restrictive in modeling nonstationarity. In this paper, we relax (2) by allowing K to depend on the location of $\mathbf{t}, \mathbf{s}$, as follows:

[R1] As $\mathbf{s} - \mathbf{t} \to 0$,

$$R(\mathbf{t}, \mathbf{s}) = K_{\mathbf{t}}(\mathbf{t} - \mathbf{s})\bigl(1 + O\bigl(|\mathbf{t} - \mathbf{s}|^{\gamma}\bigr)\bigr) \tag{3}$$

uniformly over $\mathbf{s}, \mathbf{t} \in \Omega$, with

$$K_{\mathbf{t}}(\mathbf{u}) = \Psi_{\mathbf{t}}(\mathbf{u})\mathcal{R}_{\mathbf{t}}\bigl(|\mathbf{u}|\bigr),$$

where:

(i) $\gamma \in (0, \infty)$;

(ii) $\Psi_{\mathbf{t}}(\mathbf{u})$ depends on $\mathbf{u}$ only through the direction $\mathbf{u}/|\mathbf{u}|$ and is bounded away from zero in $\mathbf{t}$ and $\mathbf{u}$; for each $\mathbf{u}$, $\Psi_{\mathbf{t}}(\mathbf{u})$ is twice-continuously differentiable in $\mathbf{t}$ and the derivatives are uniformly bounded in $\mathbf{t}$ and $\mathbf{u}$; to ensure identifiability, set $|\Psi_{\mathbf{t}}(\mathbf{u}_0)| = 1$ for some $\mathbf{u}_0$;

(iii) for each $\mathbf{t}$, $\mathcal{R}_{\mathbf{t}}(\cdot)$ is a nonnegative function which is regularly varying at 0 with index $\alpha(\mathbf{t}) \in (0, \infty)$ and such that, for any $c \in (0, \infty)$,

$$\sup_{a, |\mathbf{t}|, |\mathbf{u}| \in [0, c]} \left| \frac{\mathcal{R}_{\mathbf{t}+\varepsilon\mathbf{u}}(a\varepsilon)}{\mathcal{R}_{\mathbf{t}}(\varepsilon)} - a^{\alpha(\mathbf{t})} \right| = O\bigl(Q(\varepsilon)\bigr) \qquad \text{as } \varepsilon \to 0, \tag{4}$$

where Q only depends on c and $Q(\varepsilon) \to 0$; $\alpha(\mathbf{t})$ is twice-continuously differentiable.

We refer to [R1] as *local intrinsic stationarity*. A large class of models common in the spatial statistics literature satisfy local intrinsic stationarity. We will consider a few below. Before getting to examples, however, it might be useful to give some insight into the regular variation assumption. The simplest example of $\mathcal{R}_{\mathbf{t}}$ satisfying (iii) of [R1] is $\mathcal{R}_{\mathbf{t}}(x) = x^{\alpha(\mathbf{t})}$, $x > 0$, where α is continuously differentiable, in which case

$$\begin{aligned} &\frac{\mathcal{R}_{\mathbf{t}+\varepsilon\mathbf{u}}(a\varepsilon)}{\mathcal{R}_{\mathbf{t}}(\varepsilon)} - a^{\alpha(\mathbf{t})} \\ &\quad = a^{\alpha(\mathbf{t}+\varepsilon\mathbf{u})}\bigl(\varepsilon^{\alpha(\mathbf{t}+\varepsilon\mathbf{u})-\alpha(\mathbf{t})} - 1\bigr) + a^{\alpha(\mathbf{t})}\bigl(a^{\alpha(\mathbf{t}+\varepsilon\mathbf{u})-\alpha(\mathbf{t})} - 1\bigr) \\ &\quad = O(\varepsilon \log \varepsilon) \end{aligned} \tag{5}$$

uniformly for bounded $a, \mathbf{t}, \mathbf{u}$ and, therefore, $Q(\varepsilon) = \varepsilon \log \varepsilon$. More general examples of $\mathcal{R}_{\mathbf{t}}$ can be created by multiplying a slowly varying function to $x^{\alpha(\mathbf{t})}$ [cf. de Haan and Ferreira (2006)]. One way in which (5) is useful is that if we know

$\mathcal{R}_{\mathbf{t}}(\varepsilon)$ and $\alpha(\mathbf{t})$ then $\mathcal{R}_{\mathbf{t}+\varepsilon\mathbf{u}}(a\varepsilon)$ can be approximated by $\mathcal{R}_{\mathbf{t}}(\varepsilon)a^{\alpha(\mathbf{t})}$. We will see an application of this approximation in the inference of $K_{\mathbf{t}}$ in Section 2.2. The regular variation of $K_{\mathbf{t}}$ is related to the idea of self similarity; for instance, if $Y(\mathbf{t})$ is the standard Brownian motion on $[0, \infty)$, then, with $r = 0$ in (1), $R(\mathbf{t}, \mathbf{t}+\mathbf{u}) = |\mathbf{u}|/2$, and $Y(c\mathbf{t}) \overset{d}{=} c^{1/2}Y(\mathbf{t})$, $c > 0$.

A well-studied covariance model satisfying [R1] is the one with the expansion

$$C(\mathbf{t}, \mathbf{t}+\mathbf{h}) = c_0 + \sum_{s=1}^{k+1} c_{2s}|\mathbf{h}|^{2s} + c_{2\nu}|\mathbf{h}|^{2\nu}\big(1 + O(|\mathbf{h}|^2)\big) \qquad \text{as } \mathbf{h} \to \mathbf{0}, \tag{6}$$

where the c's are constants and $\nu \in (k, k+1)$ for some nonnegative integer k. It is easy to see that [R1] holds with $r = 2\lceil \nu \rceil$, $\alpha(\mathbf{t}) \equiv 2\nu$ and $\gamma = 2$, where $\lceil u \rceil$ denotes the smallest integer greater than or equal to u. The isotropic Matérn covariance function with smoothness parameter ν satisfies this expansion; see Stein (1999) and the references therein. In the case of the Matérn, note that the expansion in (6) contains an extra term compared with what is common in the literature. The purpose of that is to ensure $\gamma = 2$. The value of γ plays a role [cf. (12)] in the asymptotic theory of the inference problem that will be addressed in Section 2. The term $c_{2\nu}|\mathbf{h}|^{2\nu}$ is referred to as the principal irregular term in Stein (1999) and it characterizes the degree of smoothness of the sample path of Y. For the Matérn, if ν is a positive integer then the principal irregular term includes a slowly varying multiplicative factor, $\log|\mathbf{h}|$.

Suppose the covariance function of a stationary process X satisfies the expansion on the right-hand side of (6), and that F is a one-to-one and $2(k+1)$-times differentiable function from $\mathbb{R}^d$ to $\mathbb{R}^d$. Define $Y(\mathbf{t}) = X(F(\mathbf{t}))$ so that $C(\mathbf{t}, \mathbf{s}) = \operatorname{Cov}(X(F(\mathbf{t})), X(F(\mathbf{s})))$. This is the deformation process considered in Anderes and Stein (2008) and Anderes and Chatterjee (2009). These two papers have a strong influence on our work in terms of motivation and the inference approach. In Section 4.6, we show that [R1] holds with $r = \lceil 2\nu \rceil$, $\alpha(\mathbf{t}) \equiv 2\nu$, $\gamma = 1$ and $\Psi_{\mathbf{t}}(\mathbf{h}) = c_{2\nu}|J_F^{\mathbf{t}}\mathbf{h}/|\mathbf{h}||^{2\nu}$, where $J_F^{\mathbf{t}}$ is the Jacobian of F at $\mathbf{t}$. One can easily construct examples where $K_{\mathbf{t}}$ is anisotropic in this case.

Another example is the multifractional Brownian motion process on $\mathbb{R}^d$, which is a zero-mean Gaussian process with covariance function

$$C(\mathbf{t}, \mathbf{s}) = D\big(H(\mathbf{t}) + H(\mathbf{s})\big)\big\{\mathbf{t}^{H(\mathbf{t})+H(\mathbf{s})} + \mathbf{s}^{H(\mathbf{t})+H(\mathbf{s})} - |\mathbf{t}-\mathbf{s}|^{H(\mathbf{t})+H(\mathbf{s})}\big\}, \tag{7}$$

where $H(\mathbf{t})$ is Hölder continuous with range in $(0, 1)$, and

$$D(x) = \int_{\mathbb{R}^d} \frac{1 - e^{iu_1}}{|\mathbf{u}|^{x+d}}\, d\mathbf{u}, \qquad x \in (0, 1).$$

See Ayache, Shieh and Xiao (2011) and Herbin (2006) and the references therein. If $H(\mathbf{t})$ is identically equal to some constant $H \in (0, 1)$ then the process is the fractional Brownian motion introduced by Mandelbrot and Van Ness (1968). We show in Section 4.6 that if $H(\mathbf{t})$ is three-times differentiable, then [R1] holds with

$r = 2, \alpha(\mathbf{t}) = 2H(\mathbf{t}), \gamma$ equal to any constant less than 1, and $\Psi_{\mathbf{t}}(\mathbf{h}) = D(2H(\mathbf{t}))$ for all $\mathbf{h}$. We can also construct the deformation process $Y(\mathbf{t}) = X(F(\mathbf{t}))$ with X a multifractional Brownian motion, and it satisfies [R1] as well.

The structure of this paper is as follows. In Section 2, we address the inference of $K_{\mathbf{t}}$ based on differenced gridded data. Since $K_{\mathbf{t}}$ potentially changes with location, it makes sense to only use the data close to $\mathbf{t}$ when estimating $K_{\mathbf{t}}$. This is similar to the kernel estimation approach in Anderes and Stein (2008) and Anderes and Chatterjee (2009) for the deformation process. However, we adopt the local linear estimator which has a number of appealing properties. The asymptotic expressions of the variance and bias of the local linear estimator are derived for the setting where the grid becomes finer; this approach of study is sometimes referred to as fixed-domain or infill asymptotics. An interesting observation is that the differenced data are sufficiently de-correlated so that the bias and variance expressions closely match those found in the classical nonparametric regression function estimation context where there is no dependence between data points. The theoretical development of the infill asymptotics is formulated in an abstract setting, beyond local intrinsic stationarity. This will be included in Section 3. All the proofs are collected in Section 4.

We summarize the main contributions of this paper as follows. First, we introduce the notion of local intrinsic stationarity. The notion is considerably more general than that of stationarity and is satisfied by many known nonstationary models. Second, we study the inference of this model by assuming the space-varying local covariance is regularly varying and that dense data are available on a regular grid. Third, we develop an asymptotic theory for the local linear estimator of the local covariance function. Local linear (or polynomial) smoothing is in many situations the preferred nonparametric smoother but, to the best of our knowledge, has not been applied extensively in spatial statistics. By comparing with similar results in classical (nonparametric regression) settings, the rates in our asymptotic theory can be seen to be close to minimax optimal.

2. Nonparametric estimation. Assume throughout that the spatial process $\{Y(\mathbf{t}),\ t \in \Omega\}$ satisfies condition [R1]. Assume that $\Omega = [0, 1]^d$ and we observe $Y(\mathbf{t})$ for all $\mathbf{t}$ belonging to the grid

$$\mathcal{G}_n = \{(i_1, i_2, \dots, i_d)/n,\ \text{with } i_s = (j - 1/2)/n \text{ for } j = 1, \dots, n\}.$$

For convenience, the generic notation $\mathbf{t}_i = (t_{i1}, \dots, t_{id})$ will be used to denote the grid points. Below, we will analyze the asymptotic behavior of our inference procedure by allowing n to tend to ∞. Asymptotic results in the context of spatial estimation by letting the grid size to tend to zero in a fixed domain are sometimes referred to as fixed-domain or infill asymptotics.

An effective approach to decorrelate spatial data is differencing. Recursively define the differencing operators in the direction $\mathbf{h}$ of a function w on Ω by

$$\Delta_{\mathbf{h}} w(\mathbf{t}) = w(\mathbf{t}) - w(\mathbf{t} + \mathbf{h}) \quad \text{and} \quad \Delta_{\mathbf{h}}^{j} w(\mathbf{t}) = \Delta_{\mathbf{h}}^{j-1} \Delta_{\mathbf{h}} w(\mathbf{t}), \qquad j \ge 2.$$

For now, assume that $\mathcal{R}_{\mathbf{t}}$ is known. The inference of $\mathcal{R}_{\mathbf{t}}$ is important and will be discussed in Section 2.2. Consider the process $W_n(\mathbf{t},\mathbf{h})$ defined by

$$W_n(\mathbf{t},\mathbf{h}) = \{\mathcal{R}_{\mathbf{t}}(1/n)\}^{-1/2}\Delta^q_{\mathbf{h}/n}Y(\mathbf{t}) \tag{8}$$

for some integer q satisfying

$$q > \max\left(r, \frac{1}{2}\sup_{\mathbf{t}}\alpha(\mathbf{t}) + \frac{d}{2}\right). \tag{9}$$

Note that

$$\operatorname{Cov}\bigl(W_n(\mathbf{t},\mathbf{h}), W_n(\mathbf{s},\mathbf{h})\bigr) = \{\mathcal{R}_{\mathbf{t}}(1/n)\mathcal{R}_{\mathbf{s}}(1/n)\}^{-1/2}\Delta^q_{\mathbf{h}/n,\mathbf{t}}\Delta^q_{\mathbf{h}/n,\mathbf{s}}C(\mathbf{t},\mathbf{s}),$$

where $\Delta^q_{\mathbf{h}/n,\mathbf{t}}$ and $\Delta^q_{\mathbf{h}/n,\mathbf{s}}$ denote differencing with respect to $\mathbf{t}$ and $\mathbf{s}$, respectively. Thus, the assumption $q \geq r+1$ entails that the covariance of $W_n(\mathbf{t},\mathbf{h})$ does not depend on the monomials in (1). We will see in (ii) of Theorem 1 that the positive constant

$$\psi := 2q - \sup_{\mathbf{t}}\alpha(\mathbf{t}), \tag{10}$$

is a quantity that describes the covariance of $W_n(\mathbf{t},\mathbf{h})$. Note that $\psi \in (d,\infty)$ by (9).

For the inference problem considered here, we will focus on those $\mathbf{h}$ for which $W_n(\mathbf{t}_i,\mathbf{h})$ can be computed from data for all except possibly a negligible portion of the $\mathbf{t}_i$. In view of $\mathcal{G}_n$, this puts some constraint on the choice of $\mathbf{h}$. Our results can be extended with a more careful analysis for an $\mathbf{h}$ for which $W_n(\mathbf{t}_i,\mathbf{h})$ can be defined only for a portion of the data.

Define

$$g(\mathbf{u},\mathbf{t},\mathbf{h}) = \sum_{i=0}^{q}\sum_{j=0}^{q}(-1)^{i+j}\binom{q}{i}\binom{q}{j}\Psi_{\mathbf{t}}\bigl(\mathbf{u}+(i-j)\mathbf{h}\bigr)\bigl|\mathbf{u}+(i-j)\mathbf{h}\bigr|^{\alpha(\mathbf{t})}.$$

When the mean $\mu(\mathbf{t})$ of $Y(\mathbf{t})$ is sufficiently smooth, it can be shown [part (i) of Theorem 1] that

$$\mathbb{E}W_n^2(\mathbf{t},\mathbf{h}) \sim g(\mathbf{0},\mathbf{t},\mathbf{h}).$$

In that case, it is plausible to estimate $g(\mathbf{0},\mathbf{t},\mathbf{h})$ by averaging $W_n^2(\mathbf{t}_i,\mathbf{h})$ for $\mathbf{t}_i$ in a small neighborhood of $\mathbf{t}$ in some way. While it might not be immediately evident, estimating $g(\mathbf{0},\mathbf{t},\mathbf{h})$ gives us the means to estimate $K_{\mathbf{t}}$ and its components, as will be seen in Section 2.2.

The local averaging methodology adopted here for estimating $g(\mathbf{0},\mathbf{t},\mathbf{h})$ is local linear smoothing [cf. Fan and Gijbels (1996)]. Other kernel smoothing approaches could also be used, but local linear smoothing, more generally, local polynomial smoothing, is especially appealing in that the rate for the bias is the same for interior and boundary points and also the expressions for the variance and bias are tractable.

Throughout the rest of the paper, assume that k is a kernel function that satisfies the assumption:

[K] k is a nonnegative function with support $[-1,1]^d$ and has continuous first-order partial derivatives.

Let $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_d)$. For any $\mathbf{t} \in \Omega$ and bandwidth $b > 0$, let

$$\widehat{\boldsymbol{\beta}}(\mathbf{t}, \mathbf{h}; n, b) = \operatorname{argmin}_{\boldsymbol{\beta}} \sum_i k\left(\frac{\mathbf{t}_i - \mathbf{t}}{b}\right)\left\{W_n^2(\mathbf{t}_i, \mathbf{h}) - \ell(\mathbf{t}_i - \mathbf{t}; \boldsymbol{\beta})\right\}^2, \tag{11}$$

where the sum is taken on all $\mathbf{t}_i \in \mathcal{G}_n$ and

$$\ell(\mathbf{t}; \boldsymbol{\beta}) = \beta_0 + \sum_{j=1}^d \beta_j t_j \qquad \text{for } \mathbf{t} = (t_1, t_2, \ldots, t_d).$$

Then $\widehat{\boldsymbol{\beta}}(\mathbf{t}, \mathbf{h}; n, b)$ estimates

$$\boldsymbol{\beta}(\mathbf{t}, \mathbf{h}) := \left(g(\mathbf{0}, \mathbf{t}, \mathbf{h}), \frac{\partial}{\partial t_1} g(\mathbf{0}, \mathbf{t}, \mathbf{h}), \ldots, \frac{\partial}{\partial t_d} g(\mathbf{0}, \mathbf{t}, \mathbf{h})\right).$$

To obtain the large-sample behavior of $\widehat{\boldsymbol{\beta}}(\mathbf{t}, \mathbf{h}; n, b)$, it is useful to first understand the behavior of $\operatorname{Cov}(W_n(\mathbf{t}, \mathbf{h}), W_n(\mathbf{s}, \mathbf{h}))$ when $\mathbf{t}$ and $\mathbf{s}$ are close. In addition to [R1], which describes the behavior of $R(\mathbf{t}, \mathbf{s})$ in (1) when $\mathbf{t}$ and $\mathbf{s}$ are close, we also need the following complementary condition that controls the smoothness of $R(\mathbf{t}, \mathbf{s})$.

[R2] The directional partial derivative

$$f(\phi, \eta; \mathbf{t}, \mathbf{s}, \mathbf{h}) := \frac{\partial^{2q}}{\partial \phi^q \partial \eta^q} R(\mathbf{t} + \phi \mathbf{h}, \mathbf{s} + \eta \mathbf{h})$$

exists for all $\phi, \eta, \mathbf{t}, \mathbf{s}, \mathbf{h}$ such that $|(\mathbf{t} - \mathbf{s}) + (\phi - \eta)\mathbf{h}| > 0$. Further, there exist positive constants $\mathbf{c}_1$ and $\mathbf{c}_2$ such that

$$\left|f(0, 0; \mathbf{t}, \mathbf{s}, \mathbf{h})\right| \le \mathbf{c}_1 \mathcal{R}_{\mathbf{t}}\big(|\mathbf{t} - \mathbf{s}|\big)|\mathbf{t} - \mathbf{s}|^{-2q} \qquad \text{for all } \mathbf{s}, \mathbf{t} \text{ such that } |\mathbf{s} - \mathbf{t}| \in (0, \mathbf{c}_2).$$

The condition [R2] holds for the two examples, deformation process and multifractional Brownian motion, discussed in Section 1.

Define

$$\mu_n(\mathbf{t}, \mathbf{h}) = \mathbb{E} W_n(\mathbf{t}, \mathbf{h}), \qquad C_n(\mathbf{t}, \mathbf{s}, \mathbf{h}) = \operatorname{Cov}\big(W_n(\mathbf{t}, \mathbf{h}), W_n(\mathbf{s}, \mathbf{h})\big).$$

Also, let

$$\delta_n = \sup_{\mathbf{t}} \left\{\mathcal{R}_{\mathbf{t}}(1/n)\right\}^{-1/2} n^{-q}, \qquad \rho_n = Q(1/n) + n^{-\gamma \wedge 1}. \tag{12}$$

The following result addresses the behavior of the covariance of $W_n(\mathbf{u}, \mathbf{h})$ for $\mathbf{u}$ in a small neighborhood of $\mathbf{t}$.

THEOREM 1. *Assume that the conditions* [R1] *and* [R2] *hold and that q satisfies* (9). *Then we have*:

(i) *For any fixed* $\mathbf{h}$ *and* $\mathbf{u}$,

$$C_n(\mathbf{t}, \mathbf{t} + \mathbf{u}/n, \mathbf{h}) - g(\mathbf{u}, \mathbf{t}, \mathbf{h}) = O(\rho_n)$$

uniformly over $\mathbf{t} \in \Omega$.

(ii) *For any fixed* $\mathbf{h}$, *there exist positive constants* $\mathbf{c}_3$ *and* $\mathbf{c}_4$ *such that*

$$\left|C_n(\mathbf{t}, \mathbf{t} + \mathbf{u}/n, \mathbf{h})\right| \leq \mathbf{c}_4 |\mathbf{u}|^{-\psi}$$

for large n, *all* $\mathbf{t} \in \Omega$ *and all* $\mathbf{u}$ *satisfying* $2q|\mathbf{h}| < |\mathbf{u}| < \mathbf{c}_3 n$, *where* ψ *is as defined in* (10).

The rest of the paper will focus on the asymptotic properties of $\widehat{\boldsymbol{\beta}}(\mathbf{t}, \mathbf{h}; n, b)$ and related estimators. For clarity, we collect here all the assumptions that we need for that purpose, including those we have already stated. First, assume that $Y(\mathbf{t})$ is a Gaussian process satisfying [R1] and [R2] where the differencing order q satisfies (9). Further, assume that the mean function of $Y(\mathbf{t})$ is q-times continuously differentiable so that

$$\mu_n(\mathbf{t}, \mathbf{h}) = O(\delta_n)$$

with δ_n defined in (12). Finally, assume that the smoothing parameter $b = b_n$ satisfies $b \to 0$ and $nb \to \infty$.

2.1. *Asymptotic theory.* Let $\widehat{\beta}_0(\mathbf{t}, \mathbf{h}; n, b)$ denote the first element of $\widehat{\boldsymbol{\beta}}(\mathbf{t}, \mathbf{h}; n, b)$. As was explained, $\widehat{\beta}_0(\mathbf{t}, \mathbf{h}; n, b)$ is an estimate of $g(\mathbf{0}, \mathbf{t}, \mathbf{h})$. The asymptotic properties of $\widehat{\beta}_0(\mathbf{t}, \mathbf{h}; n, b)$ under the assumptions stated above are derived in Theorems 2 through 5 in Section 3, where Theorem 2 considers the bias, Theorems 3 and 4 deal with moments and asymptotic distribution, respectively, and Theorem 5 proves a uniform rate of convergence in the almost sure sense. A brief summary plus some discussions of the results are given here.

(a) In estimating $g(\mathbf{0}, \mathbf{t}, \mathbf{h})$ the bias of $\widehat{\beta}_0(\mathbf{t}, \mathbf{h}; n, b)$ is $O(b^2 + \rho_n \vee \delta_n^2)$ (Theorem 2) and the variance is $O((nb)^{-d})$ (Theorem 3). Thus, $\widehat{\beta}(\mathbf{t}, \mathbf{h}; n, b)$ is a consistent estimator of $g(\mathbf{0}, \mathbf{t}, \mathbf{h})$ for our choice of b. Both rates are uniform for all $\mathbf{t} \in \Omega$, including the boundary points.

(b) If

$$\rho_n \vee \delta_n^2 = O\big(b^2\big),$$

then the bias is $O(b^2)$ and a specific expression of the dominant term of the bias can be given in terms of the second-order derivatives of $g(\mathbf{0}, \mathbf{t}, \mathbf{h})$ in $\mathbf{t}$ and the moments of the kernel k (Theorem 2). If, additionally, $b = O(n^{-d/(d+4)})$, then $(nb)^{d/2}\{\widehat{\beta}_0(\mathbf{t}, \mathbf{h}; n, b) - g(\mathbf{0}, \mathbf{t}, \mathbf{h})\}$ is asymptotically normal (Theorem 4). These results show that the optimal estimation rate is $(n^d)^{-2/(d+4)}$, which is obtained by choosing the bandwidth $b \sim cn^{-d/(d+4)}$ for some constant c. Interestingly,

$N^{-2/(d+4)}$ is the minimax optimal rate in estimating a twice continuously differentiable function in a classical nonparametric regression setting where the sample contains N independent data points [cf. Stone (1982)]. It seems remarkable that the same rate holds in our setting where the spatial data in a fixed bounded region are highly dependent but the differencing produces the necessary level of de-correlation for the results. In this regard, similar observations were made in Kent and Wood (1997) and Chan and Wood (2000) for the stationary case.

(c) If the kernel k is of product form [K′], defined in Section 3, then for some $C < \infty$

$$\begin{aligned}\sup_{\mathbf{t}\in\Omega}&\big|\widehat{\beta}_0(\mathbf{t},\mathbf{h};n,b) - g(\mathbf{0},\mathbf{t},\mathbf{h})\big| \\ &\le C(nb)^{-d/2}(\log n)\max\{1,(nb)^{d/2}(\log n)^{-1}\big(b^2+\rho_n\vee\delta_n^2\big)\}\end{aligned} \tag{13}$$

eventually with probability one. If $\gamma > 2d/(d+4)$, then the optimal bandwidth is $b \sim cn^{-d/(d+4)}(\log n)^{2/(4+d)}$ and the optimal global rate that can be derived from (13) becomes $(n^d/\log^2 n)^{-2/(4+d)}$. This is slightly worse than the optimal global rate, $(N/\log N)^{-2/(4+d)}$, in classical nonparametric estimation [cf. Stone (1982)] with sample size N.

2.2. *Inference of $K_{\mathbf{t}}$.* So far we have assumed that $\mathcal{R}_{\mathbf{t}}$ is known in defining $\widehat{\beta}_0(\mathbf{t},\mathbf{h};n,b)$. This is typically not the case in practice. In fact, the inference of $K_{\mathbf{t}}$ is the most important goal in the inference of local intrinsic stationarity; the parameter $\alpha(\mathbf{t})$ is a quantity of interest that describes the smoothness of the surface of Y at location t while $\Psi_{\mathbf{t}}$ describes the anisotropy of the local intrinsic stationarity. This subsection discusses the estimation of $\alpha(\mathbf{t})$, $\Psi_{\mathbf{t}}$ and $\mathcal{R}_{\mathbf{t}}$. Assume that $g(\mathbf{0},\mathbf{t},\mathbf{h}) \neq 0$, which is guaranteed by $\alpha(\mathbf{t})$ not equal to an even integer.

First, one might wonder why consistent estimation of $\alpha(\mathbf{t})$ and $\Psi_{\mathbf{t}}$ can be achieved at all. To provide some intuition, let us consider a stationary Gaussian process X with mean zero and satisfying

$$\operatorname{Var}\big(\Delta_{\mathbf{h}}X(\mathbf{t})\big) = \Psi(\mathbf{h})|\mathbf{h}|^\alpha,$$

where Ψ and α does not dependent on $\mathbf{t}$. Then

$$\log\big(\Delta_{\mathbf{h}}X(\mathbf{t})\big)^2 = \log\Psi(\mathbf{h}) + \alpha\log|\mathbf{h}| + \log\chi^2,$$

where χ^2 is distributed as χ^2 with one degree of freedom. Thus, α and $\Psi(\mathbf{h})$ are the slope and intercept of a linear model. Together with the fact that increments are weakly dependent, one can be assured that α and $\Psi(\mathbf{h})$ can be consistently estimated by ordinary least squares [cf. Davies and Hall (1999)]. Although local intrinsic stationarity is considerably more general than stationarity, with modification the intuition above still applies in a local sense. We stress, however, that our estimation approach is not directly linked to this intuition.

Next, we proceed to show that the basic asymptotic theory described in Section 2.1 leads to a host of conclusions. To address the inference of $K_{\mathbf{t}}$, we begin with an idea in Chan and Wood (2000). Define

$$\xi(\mathbf{t},\mathbf{h};n,b) = \mathcal{R}_{\mathbf{t}}(1/n)\widehat{\beta}_0(\mathbf{t},\mathbf{h};n,b).$$

Note that, by (8) and (16), $\xi(\mathbf{t},\mathbf{h};n,b)$ depends only on data and not on any unknown parameters. Let

$$\widehat{\alpha}(\mathbf{t},\mathbf{h};n,b) = \frac{\log\xi(\mathbf{t},\mathbf{h};n,b) - \log\xi(\mathbf{t},\mathbf{h};2n,b)}{\log 2}.$$

Write

$$\widehat{\alpha}(\mathbf{t},\mathbf{h};n,b) - \alpha(\mathbf{t}) = A_1 + A_2,$$

where

$$A_1 = \frac{\log\mathcal{R}_{\mathbf{t}}(1/n) - \log\mathcal{R}_{\mathbf{t}}(1/(2n))}{\log 2} - \alpha(\mathbf{t})$$

and

$$A_2 = \frac{\log(1 + \frac{\widehat{\beta}_0(\mathbf{t},\mathbf{h};n,b_n) - g(\mathbf{0},\mathbf{t},\mathbf{h})}{g(\mathbf{0},\mathbf{t},\mathbf{h})}) - \log(1 + \frac{\widehat{\beta}_0(\mathbf{t},\mathbf{h};2n,b_{2n}) - g(\mathbf{0},\mathbf{t},\mathbf{h})}{g(\mathbf{0},\mathbf{t},\mathbf{h})})}{\log 2}.$$

By (4) $A_1 = O(Q(n^{-1}))$, and by the asymptotic theory for $\widehat{\beta}_0(\mathbf{t},\mathbf{h};n,b)$ we deduce

$$A_2 = O\big(\widehat{\beta}_0(\mathbf{t},\mathbf{h};n,b_n) - g(\mathbf{0},\mathbf{t},\mathbf{h}) + \widehat{\beta}_0(\mathbf{t},\mathbf{h};2n,b_{2n}) - g(\mathbf{0},\mathbf{t},\mathbf{h})\big).$$

Thus,

$$\widehat{\alpha}(\mathbf{t},\mathbf{h};n,b) - \alpha(\mathbf{t}) = O_p\big(\varepsilon_n(\mathbf{h})\big), \tag{14}$$

where $\varepsilon_n(\mathbf{h}) := \mathbb{E}^{1/2}\{\widehat{\beta}_0(\mathbf{t},\mathbf{h};n,b_n) - g(\mathbf{0},\mathbf{t},\mathbf{h})\}^2$.

Suppose that $\Psi_{\mathbf{t}}(\mathbf{h}_0) = 1$; the sign of $\Psi_{\mathbf{t}}$ is determined by $\alpha(\mathbf{t})$. Then

$$\frac{\xi(\mathbf{t},\mathbf{h};n,b)}{\xi(\mathbf{t},\mathbf{h}_0;n,b)} = \frac{\widehat{\beta}_0(\mathbf{t},\mathbf{h};n,b)}{\widehat{\beta}_0(\mathbf{t},\mathbf{h}_0;n,b)} = \Psi_{\mathbf{t}}(\mathbf{h})\big(|\mathbf{h}|/|\mathbf{h}_0|\big)^{\alpha(\mathbf{t})} + O_p\big(\varepsilon_n(\mathbf{h}) \vee \varepsilon_n(\mathbf{h}_0)\big).$$

It follows that

$$\frac{\xi(\mathbf{t},\mathbf{h};n,b)|\mathbf{h}_0|^{\widehat{\alpha}(\mathbf{t},\mathbf{h};n,b)}}{\xi(\mathbf{t},\mathbf{h}_0;n,b)|\mathbf{h}|^{\widehat{\alpha}(\mathbf{t},\mathbf{h};n,b)}} = \Psi_{\mathbf{t}}(\mathbf{h}) + O_p\big(\varepsilon_n(\mathbf{h}) \vee \varepsilon_n(\mathbf{h}_0)\big).$$

Next, for any fixed $\mathbf{u} \neq \mathbf{0}$, define

$$\widehat{K_{\mathbf{t}}(\mathbf{u}/n)} = \frac{\xi(\mathbf{t},\mathbf{h};n,b)|\mathbf{u}|^{\widehat{\alpha}(\mathbf{t},\mathbf{h};n,b)}}{\sum_{i=0}^{q}\sum_{j=0}^{q}(-1)^{i+j}\binom{q}{i}\binom{q}{j}|(i-j)\mathbf{h}|^{\widehat{\alpha}(\mathbf{t},\mathbf{h};n,b)}}.$$

It follows from (14) and the asymptotic theory for $\widehat{\beta}_0(\mathbf{t},\mathbf{h};n,b)$ that

$$\widehat{K_{\mathbf{t}}(\mathbf{u}/n)} = \Psi_{\mathbf{t}}(\mathbf{h})\mathcal{R}_{\mathbf{t}}(1/n)|\mathbf{u}|^{\alpha(\mathbf{t})}\big(1 + O_p\big(\varepsilon_n(\mathbf{h})\big)\big). \tag{15}$$

By (4),

$$
\begin{aligned}
K_{\mathbf{t}}(\mathbf{u}/n) &= \Psi_{\mathbf{t}}(\mathbf{u})\mathcal{R}_{\mathbf{t}}\big(|\mathbf{u}|/n\big) \\
&= \Psi_{\mathbf{t}}(\mathbf{u})\mathcal{R}_{\mathbf{t}}(1/n)|\mathbf{u}|^{\alpha(\mathbf{t})}\big(1 + O\big(Q\big(n^{-1}\big)\big)\big).
\end{aligned}
$$

Thus,

$$
\frac{\widehat{K_{\mathbf{t}}(\mathbf{u}/n)}}{K_{\mathbf{t}}(\mathbf{u}/n)} = 1 + O\big(Q\big(n^{-1}\big)\big) + O_p\big(\varepsilon_n(\mathbf{h})\big).
$$

Similarly, by (3) and (4),

$$
R(\mathbf{t}, \mathbf{t} + \mathbf{u}/n) = \Psi_{\mathbf{t}}(\mathbf{u})\mathcal{R}_{\mathbf{t}}(1/n)|\mathbf{u}|^{\alpha(\mathbf{t})}\big(1 + O\big(Q\big(n^{-1}\big)\big) + O\big(n^{-\gamma}\big)\big)
$$

and, therefore,

$$
\frac{\widehat{K_{\mathbf{t}}(\mathbf{u}/n)}}{R(\mathbf{t}, \mathbf{t} + \mathbf{u}/n)} = 1 + O\big(Q\big(n^{-1}\big)\big) + O\big(n^{-\gamma}\big) + O_p\big(\varepsilon_n(\mathbf{h})\big).
$$

3. A general limit theorem. We present in this section the technical results that lead to the asymptotic properties of $\widehat{\boldsymbol{\beta}}(\mathbf{t}, \mathbf{h}; n, b)$ for a fixed $\mathbf{h}$. For convenience, we will suppress $\mathbf{h}$ in the quantities $W_n(\mathbf{t}, \mathbf{h})$, $\mu_n(\mathbf{t}, \mathbf{h})$, $C_n(\mathbf{t}, \mathbf{s}, \mathbf{h})$, $g(\mathbf{u}, \mathbf{t}, \mathbf{h})$ and $\widehat{\boldsymbol{\beta}}(\mathbf{t}, \mathbf{h}; n, b)$ and refer to them as $W_n(\mathbf{t})$, $\mu_n(\mathbf{t})$, $C_n(\mathbf{t}, \mathbf{s})$, $g(\mathbf{u}, \mathbf{t})$ and $\widehat{\boldsymbol{\beta}}(\mathbf{t}; n, b)$, respectively. Then the following conditions [W1]–[W5] summarize what we know about $W_n(\mathbf{t})$ from Theorem 1 and the assumptions stated just before Section 2.1:

[W1] $W_n(\mathbf{t})$ is a Gaussian process on $\Omega = [0, 1]^d$ and is observed for all $\mathbf{t} \in \mathcal{G}_n$.

[W2] The mean function $\mu_n(\mathbf{t})$ satisfies $\mu_n(\mathbf{t}) = O(\delta_n)$ uniformly in $\mathbf{t}$ for a constant δ_n.

[W3] There exists a function $g(\mathbf{u}, \mathbf{t})$ such that $\lim_{n\to\infty} C_n(\mathbf{t}, \mathbf{t}+\mathbf{u}/n) = g(\mathbf{u}, \mathbf{t})$. The convergence is uniform for all $\mathbf{t}, \mathbf{t}+\mathbf{u}/n \in \Omega$ with $|\mathbf{u}| \le \tau$ for any given $\tau > 0$.

[W4] Define $g(\mathbf{t}) = g(\mathbf{0}, \mathbf{t})$ with $g(\mathbf{u}, \mathbf{t})$ in [W3]. $g(\mathbf{t})$ is twice continuously differentiable and there exists a constant ρ_n that tends to zero and such that $C_n(\mathbf{t}, \mathbf{t}) = g(\mathbf{t}) + O(\rho_n)$ uniformly in $\mathbf{t}$.

[W5] There exist a constant $\psi \in (d, \infty)$ and positive constants ε, τ and c_ψ such that $|C_n(\mathbf{t}, \mathbf{t}+\mathbf{u}/n)| \le c_\psi |\mathbf{u}|^{-\psi}$ for all n and $\mathbf{t}, \mathbf{t}+\mathbf{u}/n \in \Omega$ such that $\tau \le |\mathbf{u}| \le \varepsilon n$.

We will develop our asymptotic theory by focusing directly on [W1]–[W5]. This dis-association with the inference of local intrinsic stationarity potentially widens the scope of our results. From now on, $\widehat{\boldsymbol{\beta}}(\mathbf{t}; n, b)$ denotes the local linear estimator defined in (11) but with the abstract process $W_n(\mathbf{t})$ replacing $W_n(\mathbf{t}, \mathbf{h})$; accordingly, let $\widehat{\beta}_0(\mathbf{t}; n, b)$ be the first element of $\widehat{\beta}(\mathbf{t}; n, b)$. Then $\widehat{\boldsymbol{\beta}}(\mathbf{t}; n, b)$ estimates

$$
\boldsymbol{\beta}(\mathbf{t}) := \left(g(\mathbf{t}), \frac{\partial}{\partial t_1} g(\mathbf{t}), \frac{\partial}{\partial t_1} g(\mathbf{t}), \ldots, \frac{\partial}{\partial t_1} g(\mathbf{t})\right),
$$

and can be written as

$$
\widehat{\boldsymbol{\beta}}(\mathbf{t}; n, b) = \big(X' K X\big)^{-1} X' K W^2, \tag{16}
$$

where X is a matrix with rows $(1, t_{i1} - t_1, t_{i2} - t_2, \dots, t_{id} - t_d)$, K is a diagonal matrix with entries $k(\frac{\mathbf{t}_i - \mathbf{t}}{b})$, and W^2 is a column vector with entries $W_n^2(\mathbf{t}_i)$.

To address the bias $\widehat{\boldsymbol{\beta}}(\mathbf{t}; n, b)$, we first introduce some notation. Let $j_1, \dots, j_\ell$ be distinct integers in $\{1, \dots, d\}$ and $m_1, \dots, m_\ell$ be positive integers. Define

$$
\begin{aligned}
S_{j_1 \cdots j_\ell}^{m_1 \cdots m_\ell} &= S_{j_1 \cdots j_\ell}^{m_1 \cdots m_\ell}(\mathbf{t}, b) = \sum_i k\left(\frac{\mathbf{t}_i - \mathbf{t}}{b}\right)(t_{ij_1} - t_{j_1})^{m_1} \cdots (t_{ij_k} - t_{j_\ell})^{m_\ell}, \\
\kappa_{j_1 \cdots j_\ell}^{m_1 \cdots m_\ell} &= \kappa_{j_1 \cdots j_\ell}^{m_1 \cdots m_\ell}(\mathbf{t}, b) = \int_{[-\mathbf{t}/b, (1-\mathbf{t})/b]} k(\mathbf{z}) z_{j_1}^{m_1} \cdots z_{j_\ell}^{m_\ell} \, d\mathbf{z},
\end{aligned}
\tag{17}
$$

where $[-\frac{\mathbf{t}}{b}, \frac{1-\mathbf{t}}{b}] := \prod_{j=1}^d [-\frac{t_j}{b}, \frac{1-t_j}{b}]$. Note that $\kappa_{j_1 \cdots j_\ell}^{m_1 \cdots m_\ell}$ depends on $\mathbf{t}, b$ only if $\mathbf{t}$ is a "boundary point" in the sense that $\min_{1 \le j \le d} t_j \le b$ or $\max_{1 \le j \le d} t_j \ge 1 - b$. For nonboundary points, we simply have

$$
\kappa_{j_1 \cdots j_\ell}^{m_1 \cdots m_\ell} = \int k(\mathbf{z}) z_{j_1}^{m_1} \cdots z_{j_\ell}^{m_\ell} \, d\mathbf{z}.
$$

Also, write

$$
\begin{aligned}
S &= S(\mathbf{t}, b) = \sum_i k\left(\frac{\mathbf{t}_i - \mathbf{t}}{b}\right), \\
\kappa &= \kappa(\mathbf{t}, b) = \int_{[-\mathbf{t}/b, (1-\mathbf{t})/b]} k(\mathbf{z}) \, d\mathbf{z}.
\end{aligned}
\tag{18}
$$

Define

$$
\mathcal{K} = \begin{pmatrix}
\kappa & \kappa_1^1 & \kappa_2^1 & \cdots & \kappa_d^1 \\
\kappa_1^1 & \kappa_1^2 & \kappa_{12}^{11} & \cdots & \kappa_{1d}^{11} \\
\kappa_2^1 & \kappa_{12}^{11} & \kappa_2^2 & \cdots & \kappa_{2d}^{11} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\kappa_d^1 & \kappa_{1d}^{11} & \kappa_{2d}^{11} & \cdots & \kappa_d^2
\end{pmatrix}
\quad \text{and} \quad
\mathcal{N}_{i,j} = \begin{pmatrix}
\kappa_{ij}^{11} \\
\kappa_{ij1}^{111} \\
\kappa_{ij2}^{111} \\
\vdots \\
\kappa_{ijd}^{111}
\end{pmatrix}.
$$

THEOREM 2. *Assume that the assumptions* [W1]–[W4] *hold for* $\mathbf{t} \in \Omega$, *and* n *and* b *satisfy* $b \to 0$ *and* $nb \to \infty$. *Then*

$$
\begin{aligned}
\mathbb{E}\widehat{\boldsymbol{\beta}}(\mathbf{t}; n, b) = \beta(\mathbf{t}) &+ \frac{\operatorname{diag}(b^2, b, \dots, b)}{2} \mathcal{K}^{-1} \sum_{j,k=1}^d \frac{\partial^2}{\partial t_j \partial t_k} g(\mathbf{t}) \mathcal{N}_{j,k} \\
&+ \{o(1) + b^{-2} O(\rho_n \vee \delta_n^2)\}(b^2, b, \dots, b)',
\end{aligned}
\tag{19}
$$

and, *in particular*,

$$
\mathbb{E}\widehat{\beta_0}(\mathbf{t}; n, b) = g(\mathbf{t}) + \frac{b^2}{2}[\mathcal{K}^{-1}]_{1\cdot} \sum_{j,k=1}^d \frac{\partial^2}{\partial t_j \partial t_k} g(\mathbf{t}) \mathcal{N}_{j,k} + o(b^2) + O(\rho_n \vee \delta_n^2),
$$

where $[\mathcal{K}^{-1}]_{1\cdot}$ *denotes the first row of* $\mathcal{K}^{-1}$. *These results hold uniformly for* $\mathbf{t} \in \Omega$ *in the sense that the constants in the big-o and little-o terms do not depend on* $\mathbf{t}$.

Next, we consider the variance of $\widehat{\beta_0}(\mathbf{t}; n, b)$. Let

$$\widetilde{\beta}(\mathbf{t}; n, b) = \left(X'KX\right)^{-1} X'K\widetilde{W}^2,$$

where $\widetilde{W}^2$ is a vector with entries $\widetilde{W}_n^2(\mathbf{t}_i) := (W_n(\mathbf{t}_i) - \mu_n(\mathbf{t}_i))^2$. Define

$$\bar{k}(\mathbf{z}) = k(\mathbf{z})\left(\mathcal{K}_{1,1}^{-1} + \sum_{j=1}^{d} \mathcal{K}_{1,1+j}^{-1} z_j\right),$$

$$A(\mathbf{t}, b) = 2\int_{[-\mathbf{t}/b,(1-\mathbf{t})/b]} \bar{k}^2(\mathbf{z})\, d\mathbf{z} \sum_{\mathbf{j}\in\mathbb{Z}^d} g^2(\mathbf{t}, \mathbf{j}).$$

Note that, by Fatou's lemma and [W5],

$$\sum_{\mathbf{j}\in\mathbb{Z}^d} g^2(\mathbf{t}, \mathbf{j}) \le \liminf_{n\to\infty} \sum_{\mathbf{j}\in\mathbb{Z}^d} C_n^2(\mathbf{t}, \mathbf{t} + \mathbf{j}/n) \le C \sum_{j=1}^{\infty} j^{d-1-2\psi} < \infty \tag{20}$$

for some finite constant C.

Recall the definition of the "double factorial":

$$j!! = \begin{cases} j(j-2)\cdots 3\cdot 1, & \text{for odd positive integer } j,\\ j(j-2)\cdots 4\cdot 2, & \text{for even positive integer } j.\end{cases}$$

The following theorem establishes the behavior of the central moments of $\widetilde{\beta}_0(\mathbf{t}; n, b)$.

THEOREM 3. *Assume that assumptions* [W1]–[W5] *hold and that* $nb \to \infty$. *Then for* $x = 1, 2, \ldots$ *and all* $\mathbf{t} \in \Omega$,

$$\lim_{n\to\infty} (nb)^{xd/2}\mathbb{E}\left\{\frac{\widetilde{\beta}_0(\mathbf{t}; n, b) - \mathbb{E}\widetilde{\beta}_0(\mathbf{t}; n, b)}{A(\mathbf{t}, b)^{1/2}}\right\}^x = \begin{cases} 0, & \textit{for } x \textit{ odd},\\ (x-1)!!, & \textit{for } x \textit{ even}.\end{cases} \tag{21}$$

As one can expect, the proof of Theorem 3 involves computing complicated products of squared Gaussian random variables. This is done with Isserlis' theorem, or sometimes Wick's theorem or Gaussian product moment theorem.

Note that the right-hand side of (21) contains the xth moments of the standard normal distribution. Since the normal distribution is characterized by all its moments, standard weak convergence arguments entail that, for any $\mathbf{t} \in \Omega$,

$$Z_n(\mathbf{t}; n, b) := \frac{\widetilde{\beta}_0(\mathbf{t}; n, b) - \mathbb{E}\widetilde{\beta}_0(\mathbf{t}; n, b)}{\sqrt{\operatorname{Var}(\widetilde{\beta}_0(\mathbf{t}; n, b))}} \xrightarrow{d} N(0, 1) \qquad \text{as } n \to \infty \tag{22}$$

under the conclusion of Theorem 3. This together with the bias calculations in Theorem 2 lead to the asymptotic distribution of $\widehat{\beta_0}(\mathbf{t}; n, b)$.

THEOREM 4. *Assume* [W1]–[W5] *hold for* $\mathbf{t} \in \Omega$, *and that* n *and* b *satisfy* $b \to 0$ *and* $nb \to \infty$. *Then we have*

$$(nb)^{d/2} A^{-1/2}(\mathbf{t}, b)\{\widehat{\beta}_0(\mathbf{t}; n, b) - \mathbb{E}\widehat{\beta}_0(\mathbf{t}; n, b)\} \xrightarrow{d} N(0, 1) \qquad \text{as } n \to \infty, \tag{23}$$

and, *moreover*,

$$\begin{aligned} &\widehat{\beta}_0(\mathbf{t}; n, b) - g(\mathbf{t}) \\ &\quad = \frac{b^2}{2}[\mathcal{K}^{-1}]_{1\cdot} \sum_{i,j=1}^{d} \frac{\partial}{\partial t_i}\frac{\partial}{\partial t_j} g(\mathbf{t})\mathcal{N}_{i,j} + \frac{A^{1/2}(\mathbf{t}, b)}{(nb)^{d/2}} Z_n(\mathbf{t}; n, b) + R(\mathbf{t}; n, b), \end{aligned} \tag{24}$$

where

$$R(\mathbf{t}; n, b) = o(b^2) + O(\rho_n \vee \delta_n^2) + o_p((nb)^{-d/2}).$$

So far we have considered the asymptotic behavior of $\widehat{\beta}_0(\mathbf{t}; n, b)$ for a single $\mathbf{t}$. Next, we consider the global asymptotic behavior of $\widehat{\beta}_0(\mathbf{t}; n, b)$. Toward that end, we add a mild assumption that the kernel k has a "product form":

[K′] $k(\mathbf{t}) = \prod_{j=1}^{d} k_i(t_i)$ where for each i, k_i is supported on $[-1, 1]$ with $k_i(-1) = 0$ and k_i is continuously differentiable.

Let

$$V_{n,b} = \max\{1, (nb)^{d/2}(\log n)^{-1}(b^2 + \rho_n \vee \delta_n^2)\}.$$

THEOREM 5. *Assume* [W1]–[W5] *and* [K′] *hold and that* n *and* b *satisfy* $n \to \infty$ *and* $nb \to \infty$. *Then for some* $C < \infty$,

$$\sup_{\mathbf{t}\in\Omega} |\widehat{\beta}_0(\mathbf{t}; n, b) - g(\mathbf{t})| \le C(nb)^{-d/2}(\log n) V_{n,b}$$

eventually with probability 1.

4. Proofs.

4.1. *Proof of Theorem* 1. Since $\Delta_{\mathbf{h}/n}^{q}$ annihilates polynomials up to degree $m - 1$, we have

$$\begin{aligned} &C_n(\mathbf{t}, \mathbf{t} + \mathbf{u}/n, \mathbf{h}) \\ &\quad = \{\mathcal{R}_{\mathbf{t}}(n^{-1})\mathcal{R}_{\mathbf{t}+\mathbf{u}/n}(n^{-1})\}^{-1/2} \\ &\qquad \times \sum_{i=0}^{q}\sum_{j=0}^{q}(-1)^{i+j}\binom{q}{i}\binom{q}{j} R(\mathbf{t} + j\mathbf{h}/n, \mathbf{t} + \mathbf{u}/n + i\mathbf{h}/n). \end{aligned}$$

By [R1],

$$
\begin{aligned}
&\{\mathcal{R}_{\mathbf{t}}(n^{-1})\mathcal{R}_{\mathbf{t}+\mathbf{u}/n}(n^{-1})\}^{-1/2} R(\mathbf{t}+j\mathbf{h}/n, \mathbf{t}+\mathbf{u}/n+i\mathbf{h}/n)\\
&\quad = \{\mathcal{R}_{\mathbf{t}}(n^{-1})\mathcal{R}_{\mathbf{t}+\mathbf{u}/n}(n^{-1})\}^{-1/2} \Psi_{\mathbf{t}+j\mathbf{h}/n}(\mathbf{u}+(i-j)\mathbf{h})\\
&\qquad \times \mathcal{R}_{\mathbf{t}+j\mathbf{h}/n}(|\mathbf{u}+(i-j)\mathbf{h}|/n)(1+o(n^{-\gamma}))\\
&\quad = \Psi_{\mathbf{t}}(\mathbf{u}+(i-j)\mathbf{h})|\mathbf{u}+(i-j)\mathbf{h}|^{\alpha(\mathbf{t})}(1+\rho_n).
\end{aligned}
$$

This proves (i).

For (ii), we follow the proof of the lemma of Anderes and Chatterjee (2009). First, observe that we can write

$$
\Delta_{\mathbf{h}/n,1}\Delta_{\mathbf{h}/n,2}C(\mathbf{t},\mathbf{s}) = \int_0^{1/n}\int_0^{1/n} \frac{\partial^2}{\partial\phi\partial\eta} R(\mathbf{t}+\phi\mathbf{h}, \mathbf{s}+\eta\mathbf{h})\,d\eta\,d\phi,
$$

where $\Delta_{\mathbf{h},1}$, $\Delta_{\mathbf{h},2}$ are difference operator with respect to $\mathbf{t}$, $\mathbf{s}$, respectively. Generalizing this,

$$
\begin{aligned}
&\Delta^q_{\mathbf{h}/n,1}\Delta^q_{\mathbf{h}/n,2}C(\mathbf{t},\mathbf{s})\\
&\quad = \int_0^{1/n}\cdots\int_0^{1/n} \frac{\partial^{2q}}{\partial\phi_1\cdots\partial\phi_q\partial\eta_1\cdots\partial\eta_q} R\left(\mathbf{t}+\sum_{i=1}^q \phi_i\mathbf{h}, \mathbf{s}+\sum_{i=1}^q \eta_i\mathbf{h}\right) d\boldsymbol{\phi}\,d\boldsymbol{\eta} \qquad (25)\\
&\quad = \int_0^{1/n}\cdots\int_0^{1/n} f(\phi_1+\cdots+\phi_q, \eta_1+\cdots+\eta_q)\,d\boldsymbol{\phi}\,d\boldsymbol{\eta},
\end{aligned}
$$

where $f(\phi,\eta) := f(\phi,\eta;\mathbf{t},\mathbf{s},\mathbf{h}) = f(0,0;\mathbf{t}+\phi\mathbf{h}, \mathbf{s}+\eta\mathbf{h}, \mathbf{h})$, as defined in [R2]. For $\mathbf{t} = \mathbf{s}+\mathbf{u}/n$, with $\mathbf{u} = n(\mathbf{t}-\mathbf{s})$ satisfying $q|\mathbf{h}|+1 < |\mathbf{u}| < \mathbf{c}_2 n - q|\mathbf{h}|$, we have

$$
\frac{1}{n} < \frac{|\mathbf{u}|-q|\mathbf{h}|}{n} \le |\mathbf{t}'-\mathbf{s}'| < \frac{|\mathbf{u}|+q|\mathbf{h}|}{n} < \mathbf{c}_2,
$$

where $\mathbf{t}' = \mathbf{t}+\sum_{i=1}^q \phi_i\mathbf{h}$, $\mathbf{s}' = \mathbf{s}+\sum_{i=1}^q \eta_i\mathbf{h}$. Thus, [R2] implies that

$$
|f(\phi_1+\cdots+\phi_q, \eta_1+\cdots+\eta_q)| \le \mathbf{c}_1\mathcal{R}_{\mathbf{t}'}(|\mathbf{t}'-\mathbf{s}'|)|\mathbf{t}'-\mathbf{s}'|^{-2q}. \qquad (26)
$$

Now, [R1] implies that

$$
\begin{aligned}
&\{\mathcal{R}_{\mathbf{t}}(n^{-1})\mathcal{R}_{\mathbf{t}+\mathbf{u}/n}(n^{-1})\}^{-1/2}\mathcal{R}_{\mathbf{t}'}(|\mathbf{t}'-\mathbf{s}'|)|\mathbf{t}'-\mathbf{s}'|^{-2q}\\
&\qquad \sim n^{2q}\left|\mathbf{u}+n\sum_{i=1}^q(\phi_i-\eta_i)\mathbf{h}\right|^{\alpha(\mathbf{t})-2q}.
\end{aligned} \qquad (27)
$$

Combining (25)–(27), for $2q|\mathbf{h}| < |\mathbf{u}| < \mathbf{c}_3 n$,

$$
\begin{aligned}
&|C_n(\mathbf{t}, \mathbf{t}+\mathbf{u}/n, \mathbf{h})|\\
&\quad = \{\mathcal{R}_{\mathbf{t}}(n^{-1})\mathcal{R}_{\mathbf{t}+\mathbf{u}/n}(n^{-1})\}^{-1/2}\Delta^q_{\mathbf{h}/n,1}\Delta^q_{\mathbf{h}/n,2}C(\mathbf{t},\mathbf{t}+\mathbf{u}/n)\\
&\quad \le \mathbf{c}_5(|\mathbf{u}|-q|\mathbf{h}|)^{-\psi}
\end{aligned}
$$

for some finite $\mathbf{c}_5$. Thus, (ii) follows by adjusting the constant $\mathbf{c}_5$.

4.2. *Proof of Theorem 2*.

LEMMA 1. *For any nonnegative integers $m_1, \dots, m_d$,*

$$\left|n^{-d} b^{-d-\sum_{\ell=1}^d m_\ell} \sum_i k\left(\frac{\mathbf{t}_i - \mathbf{t}}{b}\right) \prod_{\ell=1}^d t_{\mathbf{i}\ell}^{m_\ell} - \int_{[-\mathbf{t}/b,(1-\mathbf{t})/b]} k(\mathbf{z}) \prod_{\ell=1}^d z_\ell^{m_\ell}\, d\mathbf{z}\right| \\ \le cn^{-1} b^{-1},$$

for some finite constant c, uniformly for $\mathbf{t} \in \Omega$.

PROOF. Let $\mathcal{C}_i$ denote the d-dimensional cube with area n^{-d} centered at $\mathbf{t}_i$ and $f(\mathbf{z}) = k(\mathbf{z}) \prod_{\ell=1}^d z_\ell^{m_\ell}$. Changing variables with $\mathbf{z} = (\mathbf{s} - \mathbf{t})/b$,

$$b^{-d} \sum_i \int_{\mathcal{C}_i} f\left(\frac{\mathbf{s}-\mathbf{t}}{b}\right) d\mathbf{s} = b^{-d} \int_{[0,1]^d} f\left(\frac{\mathbf{s}-\mathbf{t}}{b}\right) d\mathbf{s} = \int_{[-\mathbf{t}/b,(1-\mathbf{t})/b]} f(\mathbf{z})\, d\mathbf{z}.$$

Thus,

$$\left|n^{-d} b^{-d-\sum_{\ell=1}^d m_\ell} \sum_i k\left(\frac{\mathbf{t}_i - \mathbf{t}}{b}\right) \prod_{\ell=1}^d t_{i\ell}^{m_\ell} - \int_{[-\mathbf{t}/b,(1-\mathbf{t})/b]} f(\mathbf{z})\, d\mathbf{z}\right| \\ = b^{-d}\left|n^{-d} \sum_i f\left(\frac{\mathbf{t}_i - \mathbf{t}}{b}\right) - \sum_i \int_{S_i} f\left(\frac{\mathbf{s}-\mathbf{t}}{b}\right) d\mathbf{s}\right|.$$

By the mean value theorem,

$$\left|f\left(\frac{\mathbf{t}_i - \mathbf{t}}{b}\right) - f\left(\frac{\mathbf{s}-\mathbf{t}}{b}\right)\right| = \left|\left(\frac{\mathbf{s}-\mathbf{t}_i}{b}\right)' (\nabla f)\left(\frac{\mathbf{z}-\mathbf{t}}{b}\right)\right|$$

for some $\mathbf{z}$ between $\mathbf{s}$ and $\mathbf{t}_i$, where ∇f is the gradient of f. For $\mathbf{s} \in \mathcal{C}_i$, $|\frac{\mathbf{s}-\mathbf{t}_i}{b}| \le n^{-1} b^{-1}$, and hence

$$b^{-d}\left|n^{-d} \sum_i f\left(\frac{\mathbf{t}_i - \mathbf{t}}{b}\right) - \sum_i \int_{S_i} f\left(\frac{\mathbf{s}-\mathbf{t}}{b}\right) d\mathbf{s}\right| \le n^{-1} b^{-1} \|\nabla f\|_\infty,$$

where $\|\nabla f\|_\infty = \max_{\mathbf{z} \in [-1,1]^d} |(\nabla f)(\mathbf{z})|$. $\square$

PROOF OF THEOREM 2. By Taylor's theorem and the assumptions [W1]–[W4],

$$\begin{aligned} \mathbb{E}\{W_n^2(\mathbf{t}_i)\} &= C_n(\mathbf{t}_i, \mathbf{t}_i) + \mu^2(\mathbf{t}_i) \\ &= g(\mathbf{t}) + (\mathbf{t}_i - \mathbf{t})' \nabla g(\mathbf{t}) + \tfrac{1}{2} (\mathbf{t}_i - \mathbf{t})' H_g(\mathbf{t}) (\mathbf{t}_i - \mathbf{t}) \\ &\quad + o(|\mathbf{t}_i - \mathbf{t}|^2) + O(\rho_n \wedge \delta_n^2) \end{aligned} \tag{28}$$

uniformly for $\mathbf{t} \in [0, 1]^d$, where H_g is the Hessian of g. Define

$$M_0 = \begin{pmatrix} S \\ S_1^1 \\ S_2^1 \\ \vdots \\ S_d^1 \end{pmatrix}, \qquad M_i = \begin{pmatrix} S_i^1 \\ S_{i1}^{11} \\ S_{i2}^{11} \\ \vdots \\ S_{id}^{11} \end{pmatrix}, \qquad N_{i,j} = \begin{pmatrix} S_{ij}^{11} \\ S_{ij1}^{111} \\ S_{ij2}^{111} \\ \vdots \\ S_{ijd}^{111} \end{pmatrix}, \qquad \mathcal{N}_{i,j} = \begin{pmatrix} \kappa_{ij}^{11} \\ \kappa_{ij1}^{111} \\ \kappa_{ij2}^{111} \\ \vdots \\ \kappa_{ijd}^{111} \end{pmatrix},$$

where we use the notation defined in (17) and (18). It then follows from (28) that

$$\begin{aligned} &X'K\mathbb{E}(W^2) \\ &\quad = g(\mathbf{t})M_0 + \sum_{i=1}^d \frac{\partial}{\partial t_i} g(\mathbf{t}) M_i + \frac{1}{2}\sum_{i,j=1}^d \frac{\partial}{\partial t_i}\frac{\partial}{\partial t_j} g(\mathbf{t}) N_{i,j} + R, \end{aligned} \tag{29}$$

where

$$R = \begin{pmatrix} \sum_i k\left(\frac{\mathbf{t}_i - \mathbf{t}}{b}\right)\{o(|\mathbf{t}_i - \mathbf{t}|^2) + O(\rho_n \wedge \delta_n^2)\} \\ \sum_i k\left(\frac{\mathbf{t}_i - \mathbf{t}}{b}\right) t_{i1}\{o(|\mathbf{t}_i - \mathbf{t}|^2) + O(\rho_n \wedge \delta_n^2)\} \\ \vdots \\ \sum_i k\left(\frac{\mathbf{t}_i - \mathbf{t}}{b}\right) t_{id}\{o(|\mathbf{t}_i - \mathbf{t}|^2) + O(\rho_n \wedge \delta_n^2)\} \end{pmatrix}. \tag{30}$$

Note that M_ℓ, $0 \le \ell \le d$, are the columns of $X'KX$. Thus, $(X'KX)^{-1}X'KX = I_{d+1}$ and we have

$$(X'KX)^{-1} M_\ell = e_{\ell+1}, \qquad 0 \le \ell \le d, \tag{31}$$

where e_j is a column vector with a 1 in the jth row and zeros elsewhere. Combining (29) and (31),

$$(X'KX)^{-1}\left(g(\mathbf{t})M_0 + \sum_{i=1}^d \frac{\partial}{\partial t_i} g(\mathbf{t}) M_i\right) = \beta(\mathbf{t}). \tag{32}$$

By Lemma 1, if we let $D = \operatorname{diag}(1, b, \dots, b)$,

$$\begin{aligned} (X'KX)^{-1} &= [n^d b^d D(\mathcal{K} + O(n^{-1}b^{-1}))D]^{-1} \\ &= n^{-d} b^{-d} D^{-1}(\mathcal{K}^{-1} + O(n^{-1}b^{-1}))D^{-1}. \end{aligned} \tag{33}$$

Also, by Lemma 1,

$$
\begin{aligned}
(34)\qquad & \sum_{i,j=1}^{d} \frac{\partial}{\partial t_i}\frac{\partial}{\partial t_j} g(\mathbf{t}) N_{i,j} \\
&= n^d b^{d+2} D \sum_{i,j=1}^{d} \frac{\partial}{\partial t_i}\frac{\partial}{\partial t_j} g(\mathbf{t}) \mathcal{N}_{i,j} + (nb)^{-1} O \begin{pmatrix} n^d b^{d+2} \\ n^d b^{d+3} \\ \vdots \\ n^d b^{d+3} \end{pmatrix}.
\end{aligned}
$$

Combining (33) and (34),

$$
\begin{aligned}
(35)\qquad & (X'KX)^{-1} \sum_{i,j=1}^{d} \frac{\partial}{\partial t_i}\frac{\partial}{\partial t_j} g(\mathbf{t}) N_{i,j} \\
&= b^2 D^{-1} \mathcal{K}^{-1} \sum_{i,j=1}^{d} \frac{\partial}{\partial t_i}\frac{\partial}{\partial t_j} g(\mathbf{t}) \mathcal{N}_{i,j} + O(n^{-1} b) D^{-1} \mathbf{1}.
\end{aligned}
$$

By (33) and (30),

$$
(36)\qquad (X'KX)^{-1} R = \begin{pmatrix} o(b^2) + O(\rho_n \wedge \delta_n^2) \\ o(b) + b^{-1} O(\rho_n \wedge \delta_n^2) \\ \vdots \\ o(b) + b^{-1} O(\rho_n \wedge \delta_n^2) \end{pmatrix}.
$$

Thus, we obtain (19) by combining (32), (35) and (36), where the remainder term in (35) can be absorbed by other terms using the assumptions on b and n. $\square$

4.3. *Proof of Theorem* 3. We first discuss a result, known as Isserlis' theorem or Wick's theorem, that can be applied to compute higher-order moments of the multivariate normal distribution in terms of its covariance matrix.

THEOREM 6. *Suppose m, n are nonnegative integers and at least one is nonzero, and $X_1, X_2, \dots, X_m, X_{m+1}, \dots, X_{m+n}$ are jointly Gaussian with mean* 0. *Then for n odd, we have*

$$
\mathbb{E}\left(\prod_{i=1}^{m} (X_i^2 - \mathbb{E}(X_i^2)) \cdot \prod_{j=1}^{n} X_{m+j} \right) = 0
$$

and for n even,

$$
\begin{aligned}
& \mathbb{E}\left(\prod_{i=1}^{m} (X_i^2 - \mathbb{E}(X_i^2)) \cdot \prod_{j=1}^{n} X_{m+j} \right) \\
&\quad = \sum_{(i_1,i_2),\dots,(i_{2m-1},i_{2m}),(i_{2m+1},i_{2m+2}),\dots,(i_{2m+n-1},i_{2m+n})} \mathbb{E}(X_{i_1} X_{i_2}) \cdots \\
&\qquad \times \mathbb{E}(X_{i_{2m+n-1}} X_{i_{2m+n}}),
\end{aligned}
$$

where the sum is over all unique pairwise samplings from $\{1, 1, 2, 2, \dots, m-1, m-1, m, m, m+1, m+2, \dots, m+n\}$ *such that no pairing has identical elements. In particular,*

$$\mathbb{E}\left(\prod_{i=1}^{m}(X_i^2 - \mathbb{E}(X_i^2))\right) = \sum_{(i_1,i_2),\dots,(i_{2m-1},i_{2m})} \mathbb{E}(X_{i_1}X_{i_2})\cdots\mathbb{E}(X_{i_{2m-1}}X_{i_{2m}}),$$

where the sum is over all unique pairwise samplings from $\{1, 1, 2, 2, \dots, m-1, m-1, m, m\}$ *such that no pairing has identical elements.*

PROOF. The proof is by induction on m. For $m = 0$ and arbitrary n, the result is known as Isserlis' or Wick's theorem [cf. Isserlis (1918)]. Suppose the formulas hold for $0, 1, 2, \dots, m$ and arbitrary n. Then

$$\begin{aligned}
&\mathbb{E}\left(\prod_{i=1}^{m+1}(X_i^2 - \mathbb{E}(X_i^2)) \cdot \prod_{j=1}^{n} X_{m+1+j}\right) \\
&\quad = \mathbb{E}\left(\prod_{i=1}^{m}(X_i^2 - \mathbb{E}(X_i^2)) \cdot X_{m+1}^2 \prod_{j=1}^{n} X_{m+1+j}\right) \\
&\qquad - \mathbb{E}(X_{m+1}^2) \cdot \mathbb{E}\left(\prod_{i=1}^{m}(X_i^2 - \mathbb{E}(X_i^2)) \cdot \prod_{j=1}^{n} X_{m+1+j}\right)
\end{aligned}$$

and by the induction hypothesis, the above is 0 if n is odd and is equal to the desired expression if n is even. $\square$

PROOF OF THEOREM 3. By (33), for any $\mathbf{v} \in \mathbb{R}^{n^d}$, we can write

$$\begin{aligned}
&[(X'KX)^{-1}X'K\mathbf{v}]_1 \\
&\quad = (nb)^{-d}\sum_i k\left(\frac{\mathbf{t}_i - \mathbf{t}}{b}\right)\left\{\mathcal{K}_{1,1}^{-1} + \sum_{j=1}^{d}\mathcal{K}_{1,j+1}^{-1}\left(\frac{t_{ij} - t_j}{b}\right) + O((nb)^{-1})\right\}v_i \\
&\quad =: (nb)^{-d}\sum_i \breve{k}\left(\frac{\mathbf{t}_i - \mathbf{t}}{b}\right)v_i.
\end{aligned}$$

Note that $\breve{k}$ is uniformly bounded. Let

$$\widetilde{W}_n(\mathbf{t}) = W_n(\mathbf{t}) - \mu_n(\mathbf{t}),$$

and $\widetilde{W}^2$ denote the vector with elements $\widetilde{W}_n(\mathbf{t}_i)^2$ and let $\tilde{\beta}(\mathbf{t})$ be defined with $W_n(\mathbf{t}_i)$ replaced by $\widetilde{W}_n(\mathbf{t}_i)$. Then we can write

$$\begin{aligned}
&\tilde{\beta}_0(\mathbf{t}) - \mathbb{E}\tilde{\beta}_0(\mathbf{t}) \\
&\quad = [(X'KX)^{-1}X'K(\widetilde{W}^2 - \mathbb{E}\widetilde{W}^2)]_1 \qquad (37)\\
&\quad = (nb)^{-d}\sum_i \breve{k}\left(\frac{\mathbf{t}_i - \mathbf{t}}{b}\right)\{\widetilde{W}_n^2(\mathbf{t}_i) - \mathbb{E}\widetilde{W}_n^2(\mathbf{t}_i)\}.
\end{aligned}$$

In this result, $\mathbf{t}$ is an arbitrary point in Ω. However, in the proof below we will focus on the case where $\mathbf{t}$ is a grid point in $\mathcal{G}_n$. If $\mathbf{t}$ is a nongrid point, then we can work with the closest grid point and the bias can be absorbed by the rest of the terms.

First, we will show that (21) holds for $x = 2$. Applying Theorem 6,

$$\mathbb{E}\{\widetilde{\beta}_0(\mathbf{t}) - \mathbb{E}(\widetilde{\beta}_0(\mathbf{t}))\}^2 = 2(nb)^{-d} \sum_{i,j} \breve{k}\left(\frac{\mathbf{t}_i - \mathbf{t}}{b}\right)\breve{k}\left(\frac{\mathbf{t}_j - \mathbf{t}}{b}\right) C_n^2(\mathbf{t}_i, \mathbf{t}_j). \tag{38}$$

Relabel the points $\mathbf{t}_i$ as $\mathbf{t} + \mathbf{i}/n$ with $\mathbf{i}$ in some subset of $\mathbb{Z}^d$. For a fixed positive m, write

$$\begin{aligned}
&(nb)^{-d} \sum_{\mathbf{i},\mathbf{j}} \breve{k}\left(\frac{\mathbf{i}}{nb}\right)\breve{k}\left(\frac{\mathbf{j}}{nb}\right) C_n^2(\mathbf{t} + \mathbf{i}/n, \mathbf{t} + \mathbf{j}/n) \\
&\quad - \int_{[-\mathbf{t}/b, (1-\mathbf{t})/b]} \bar{k}^2(\mathbf{z})\, d\mathbf{z} \sum_{\mathbf{j} \in \mathbb{Z}^d} g^2(\mathbf{t}, \mathbf{j}) \\
&\quad = \varepsilon_{n,m,1} + \varepsilon_{n,m,2} + \varepsilon_{n,m,3} + \varepsilon_{n,m,4} + \varepsilon_{n,m,5},
\end{aligned}$$

where

$$\begin{aligned}
\varepsilon_{n,m,1} &= (nb)^{-d} \left(\sum_{\mathbf{i},\mathbf{j}} - \sum_{\mathbf{i}} \sum_{\mathbf{j}: |\mathbf{j}-\mathbf{i}| \le m} \right) \breve{k}\left(\frac{\mathbf{i}}{nb}\right)\breve{k}\left(\frac{\mathbf{j}}{nb}\right) C_n^2(\mathbf{t} + \mathbf{i}/n, \mathbf{t} + \mathbf{j}/n), \\
\varepsilon_{n,m,2} &= (nb)^{-d} \left(\sum_{\mathbf{i}} \sum_{|\mathbf{j}| \le m} \breve{k}\left(\frac{\mathbf{i}}{nb}\right)\breve{k}\left(\frac{\mathbf{i}+\mathbf{j}}{nb}\right) C_n^2(\mathbf{t} + \mathbf{i}/n, \mathbf{t} + (\mathbf{i}+\mathbf{j})/n) \right. \\
&\qquad \left. - \sum_{\mathbf{i}} \breve{k}^2\left(\frac{\mathbf{i}}{nb}\right) \sum_{|\mathbf{j}| \le m} C_n^2\big(\mathbf{t} + \mathbf{i}/n, \mathbf{t} + (\mathbf{i}+\mathbf{j})/n\big) \right), \\
\varepsilon_{n,m,3} &= (nb)^{-d} \sum_{\mathbf{i}} \breve{k}^2\left(\frac{\mathbf{i}}{nb}\right) \sum_{|\mathbf{j}| \le m} C_n^2\big(\mathbf{t} + \mathbf{i}/n, \mathbf{t} + (\mathbf{i}+\mathbf{j})/n\big) \\
&\qquad - \int_{[-\mathbf{t}/b, (1-\mathbf{t})/b]} \bar{k}^2(\mathbf{z})\, d\mathbf{z} \sum_{|\mathbf{j}| \le m} C_n^2\big(\mathbf{t} + \mathbf{i}/n, \mathbf{t} + (\mathbf{i}+\mathbf{j})/n\big), \\
\varepsilon_{n,m,4} &= \int_{[-\mathbf{t}/b, (1-\mathbf{t})/b]} \bar{k}^2(\mathbf{z})\, d\mathbf{z} \sum_{|\mathbf{j}| \le m} C_n^2\big(\mathbf{t} + \mathbf{i}/n, \mathbf{t} + (\mathbf{i}+\mathbf{j})/n\big) \\
&\qquad - \int_{[-\mathbf{t}/b, (1-\mathbf{t})/b]} \bar{k}^2(\mathbf{z})\, d\mathbf{z} \sum_{|\mathbf{j}| \le m} g(\mathbf{t}, \mathbf{j})^2, \\
\varepsilon_{n,m,5} &= \int_{[-\mathbf{t}/b, (1-\mathbf{t})/b]} \bar{k}^2(\mathbf{z})\, d\mathbf{z} \sum_{|\mathbf{j}| \le m} g^2(\mathbf{t}, \mathbf{j}) - \int_{[-\mathbf{t}/b, (1-\mathbf{t})/b]} \bar{k}^2(\mathbf{z})\, d\mathbf{z} \sum_{\mathbf{j} \in \mathbb{Z}^d} g^2(\mathbf{t}, \mathbf{j}).
\end{aligned}$$

We need to show that, for $i = 1, 2, \dots, 5$,

$$\lim_{m\to\infty} \limsup_{n\to\infty} \varepsilon_{n,m,i} = 0. \tag{39}$$

First,

$$\varepsilon_{n,m,1} = (nb)^{-d} \sum_{\mathbf{i}} \sum_{|\mathbf{j}|>m} \breve{k}\left(\frac{\mathbf{i}}{nb}\right)\breve{k}\left(\frac{\mathbf{i}+\mathbf{j}}{nb}\right) C_n^2(\mathbf{t}+\mathbf{i}/n, \mathbf{t}+(\mathbf{i}+\mathbf{j})/n).$$

Take $m > \tau$. By [W5],

$$\sup_n \sup_{\mathbf{j}:|\mathbf{j}|>m} \frac{C_n^2(\mathbf{t}+\mathbf{i}/n, \mathbf{t}+(\mathbf{i}+\mathbf{j})/n)}{|\mathbf{j}|^{-2\psi}} \le c_\psi^2. \tag{40}$$

Thus,

$$\varepsilon_{n,m,1} \le c_\psi^2 (nb)^{-d} \sum_{\mathbf{i}} \breve{k}\left(\frac{\mathbf{i}}{nb}\right) \sum_{|\mathbf{j}|>m} \breve{k}\left(\frac{\mathbf{i}+\mathbf{j}}{nb}\right)|\mathbf{j}|^{-2\psi}.$$

We conclude that (39) holds for $i = 1$ using the facts that $\breve{k}$ is bounded and supported on $[-1, 1]^d$, and $\sum_{|\mathbf{j}|>m} |\mathbf{j}|^{-2\psi} = O(m^{d-1-2\psi}) \to 0$ as $m \to \infty$.

Next, since $C_n(\mathbf{s}, \mathbf{t})$, k and k' are bounded,

$$\begin{aligned}
&|\varepsilon_{n,m,2}| \\
&\quad \le (nb)^{-d} \sum_{\mathbf{i}} \sum_{|\mathbf{j}|\le m} \breve{k}\left(\frac{\mathbf{i}}{nb}\right)\left|\breve{k}\left(\frac{\mathbf{i}+\mathbf{j}}{nb}\right) - \breve{k}\left(\frac{\mathbf{i}}{nb}\right)\right| C_n^2(\mathbf{t}+\mathbf{i}/n, \mathbf{t}+(\mathbf{i}+\mathbf{j})/n) \\
&\quad = O\big((nb)^{-1}\big),
\end{aligned}$$

which shows (39) for $i = 2$.

By Riemann approximation and the fact that $\breve{k}$ and $\bar{k}$ differ by $O((nb)^{-1})$ uniformly, we conclude that (39) holds for $i = 3$. Finally, that (39) holds for $i = 4$ and 5 follow from [W3] and (20), respectively. Thus, the proof of (21) for $x = 2$ is accomplished.

To show (21) for a general $x > 2$, let $\delta_{ij} = \breve{k}(\frac{\mathbf{t}_i-\mathbf{t}}{b})^{1/2}\breve{k}(\frac{\mathbf{t}_j-\mathbf{t}}{b})^{1/2} C_n(\mathbf{t}_i, \mathbf{t}_j)$. By Theorem 6, we can write

$$\begin{aligned}
&(nb)^{xd/2}\mathbb{E}\{\widetilde{\beta}_0(\mathbf{t}; n, b) - \mathbb{E}(\widetilde{\beta}_0(\mathbf{t}; n, b))\}^x \\
&\qquad = (nb)^{-xd/2} \sum_{i_1,\dots,i_x} \sum_{\mathbf{S}_x} \delta_{s_1 s_2}\delta_{s_3 s_4}\cdots\delta_{s_{2x-1}s_{2x}},
\end{aligned} \tag{41}$$

where $\mathbf{S}_x$ is the set of all possible ways to make x pairs, $\{(s_1, s_2), \dots, (s_{2x-1}, s_{2x})\}$, with s_j chosen from $\mathbf{I}_x = \{i_1, i_1, i_2, i_2, \dots, i_x, i_x\}$ without replacement and members within each pair must different.

To explain the ideas of the general proof, we first consider the cases $x = 3$ and 4. For $x = 3$, for any given indices i_1, i_2, i_3, the number of all possible pairings in

$\mathbf{S}_x$ is 8, and each pairing can be expressed as $\{(j_1, j_2), (j_2, j_3), (j_3, j_1)\}$, where (j_1, j_2, j_3) is a permutation of (i_1, i_2, i_3). Thus,

$$\begin{aligned}
&\Big|\sum_{i_1,i_2,i_3}\sum_{\mathbf{S}_x}\delta_{s_1s_2}\delta_{s_3s_4}\delta_{s_5s_6}\Big| \\
&\qquad = 8\Big|\sum_{i_1,i_2,i_3}\delta_{i_1i_2}\delta_{i_1i_3}\delta_{i_2i_3}\Big| \\
&\qquad \le 8\sum_{i_1,i_2}|\delta_{i_1i_2}|\frac{1}{2}\Big(\sum_{i_3}\delta^2_{i_1i_3}+\sum_{i_3}\delta^2_{i_2i_3}\Big) \\
&\qquad \le 8\mathbf{c}_6\mathbf{c}_7(nb)^d,
\end{aligned} \tag{42}$$

where $\mathbf{c}_6$ and $\mathbf{c}_7$ are finite bounds for $\sum_{i_2}|\delta_{i_1i_2}|$ and $\sum_{i_2}\delta^2_{i_1i_2}$, respectively, which are guaranteed finite by [W5].

For $x = 4$, $\mathbf{S}_x$ contains only two possible configurations:

$$\{(j_1, j_2), (j_2, j_3), (j_3, j_4), (j_4, j_1)\} \quad \text{or} \quad \{(j_1, j_2), (j_2, j_1), (j_3, j_4), (j_4, j_3)\},$$

where (j_1, j_2, j_3, j_4) is a permutation of (i_1, i_2, i_3, i_4). Separating the two cases,

$$\begin{aligned}
&\sum_{i_1,\dots,i_4}\sum_{\mathbf{S}_x}\delta_{s_1s_2}\delta_{s_3s_4}\cdots\delta_{s_7s_8} \\
&\qquad = 6\sum_{i_1,\dots,i_4}\delta_{i_1i_2}\delta_{i_2i_3}\delta_{i_3i_4}\delta_{i_4i_1} + 12\sum_{i_1,i_2}\delta^2_{i_1i_2}\sum_{i_3,i_4}\delta^2_{i_3i_4}.
\end{aligned} \tag{43}$$

Similar to (42) the first term on the right is bounded by

$$6\sum_{i_1,i_2,i_3}|\delta_{i_1i_2}\delta_{i_2i_3}|\frac{1}{2}\Big(\sum_{i_4}\delta^2_{i_3i_4}+\sum_{i_4}\delta^2_{i_4i_1}\Big) \le 6\mathbf{c}_6^2\mathbf{c}_7(nb)^d. \tag{44}$$

The second term on the right-hand side of (43) is $3A^2(\mathbf{t}, b) + O(n^{-1}b^{-1})$ by proof for the case $x = 2$ above.

Now consider any general $x \ge 3$. Since each index i_j appears exactly twice, each pairing $\{(s_1, s_2), (s_3, s_4), \dots, (s_{2x-1}, s_{2x})\}$ can be partitioned into a collection of subsets, "chains" of the form $\{(j_1, j_2), (j_2, j_3), \dots, (j_\ell, j_1)\}$. For convenience, we say the chain $\{(j_1, j_2), (j_2, j_3), \dots, (j_\ell, j_1)\}$ has length $\ell \ge 2$. Then, similar to (42) and (44),

$$\begin{aligned}
&\Big|\sum_{j_1,\dots,j_\ell}\delta_{j_1j_2}\cdots\delta_{j_\ell j_1}\Big| \\
&\qquad \le \sum_{j_1,\dots,j_{\ell-1}}|\delta_{j_1j_2}\cdots\delta_{j_{\ell-2}j_{\ell-1}}|\frac{1}{2}\Big(\sum_{j_\ell}\delta^2_{j_{\ell-1}j_\ell}+\sum_{j_\ell}\delta^2_{j_\ell j_1}\Big) \\
&\qquad \le \mathbf{c}_6^{\ell-2}\mathbf{c}_7(nb)^d \\
&\qquad \le \mathbf{c}_8^{\ell}(nb)^d,
\end{aligned} \tag{45}$$

where $\mathbf{c}_8 = \max(1, \mathbf{c}_6, \mathbf{c}_7)$.

For a given pairing $\{(s_1, s_2), (s_3, s_4), \ldots, (s_{2x-1}, s_{2x})\}$, suppose the partition comprises of m chains with lengths $\ell_1, \ell_2, \ldots, \ell_m$, where $\ell_1 + \cdots + \ell_m = x$. Then from (45),

$$\left|\sum_{i_1,\ldots,i_x} \delta_{s_1 s_2}\delta_{s_3 s_4}\cdots\delta_{s_{2x-1}s_{2x}}\right| \le \mathbf{c}_8^x (nb)^{md}$$

and we have

$$(nb)^{-xd/2}\left|\sum_{i_1,\ldots,i_x} \delta_{s_1 s_2}\delta_{s_3 s_4}\cdots\delta_{s_{2x-1}s_{2x}}\right| \le \mathbf{c}_8^x (nb)^{-(x-2m)d/2}. \tag{46}$$

If $x > 2m$, then this expression tends to 0. If x is even and $m = x/2$, then this partition will contain $x/2$ chains of length 2 and, therefore,

$$\begin{aligned}
(nb)^{-xd/2} &\sum_{i_1,\ldots,i_x} \delta_{s_1 s_2}\delta_{s_3 s_4}\cdots\delta_{s_{2x-1}s_{2x}} \\
&= (nb)^{-xd/2}\left(\sum_{i_1,i_2} \delta_{s_1 s_2}^2\right)^{x/2} \\
&= 2^{-x/2}\big(A(\mathbf{t})\big)^{x/2} + o(1).
\end{aligned}$$

Since the number of ways to obtain $x/2$ chains of length two from $\mathbf{I}_x$ is $(2(x-1))\cdot(2(x-3))\cdots 2 = (x-1)!!2^{x/2}$,

$$(nb)^{-xd/2} \sum_{i_1,\ldots,i_x} \delta_{s_1 s_2}\delta_{s_3 s_4}\cdots\delta_{s_{2x-1}s_{2x}} = (x-1)!!\big(A(\mathbf{t})\big)^{x/2} + o(1).$$

If x is odd, then m cannot be equal to $x/2$ and, therefore,

$$(nb)^{-xd/2} \sum_{i_1,\ldots,i_k} \sum_{\mathbf{S}_x} \delta_{s_1 s_2}\delta_{s_3 s_4}\cdots\delta_{s_{2x-1}s_{2x}} \to 0.$$

□

4.4. *Proof of Theorem* 4. Since

$$W_n^2(\mathbf{t}) - \mathbb{E}W_n^2(\mathbf{t}) = \{\widetilde{W}^2(\mathbf{t}) - \mathbb{E}\widetilde{W}^2(\mathbf{t})\} + 2\mu_n(\mathbf{t})\big(W_n(\mathbf{t}) - \mu_n(\mathbf{t})\big),$$

by (37) we have

$$\widehat{\beta}_0(\mathbf{t}) - \mathbb{E}\widehat{\beta}_0(\mathbf{t}) = \{\widetilde{\beta}_0(\mathbf{t}) - \mathbb{E}\widetilde{\beta}_0(\mathbf{t})\} + U_n(\mathbf{t}),$$

where

$$U_n(\mathbf{t}) = 2(nb)^{-d}\sum_i \breve{k}\left(\frac{\mathbf{t}_i - \mathbf{t}}{b}\right)\mu_n(\mathbf{t}_i)\{W_n(\mathbf{t}_i) - \mu_n(\mathbf{t}_i)\}. \tag{47}$$

Clearly, $U_n(\mathbf{t})$ has normal distribution with mean zero and variance

$$
\begin{aligned}
4(nb)^{-2d}\sum_i\sum_j \breve{k}\Big(\frac{\mathbf{t}_i-\mathbf{t}}{b}\Big)\breve{k}\Big(\frac{\mathbf{t}_j-\mathbf{t}}{b}\Big)\mu_n(\mathbf{t}_i)\mu_n(\mathbf{t}_j)C_n(\mathbf{t}_i,\mathbf{t}_j)&\\
=O\big((nb)^{-d}\delta_n^2\big),&
\end{aligned}
$$

and so

$$
U_n(\mathbf{t})=O_p\big((nb)^{-d/2}\delta_n\big). \tag{48}
$$

Thus, $(nb)^{d/2}U_n(\mathbf{t})=o_p(1)$ and (22) implies (23). To obtain (24), write

$$
\begin{aligned}
\widehat{\beta}_0(\mathbf{t})-g(\mathbf{t})&\\
&=\{\mathbb{E}\widehat{\beta}_0(\mathbf{t})-g(\mathbf{t})\}+\{\widehat{\beta}_0(\mathbf{t})-\mathbb{E}\widehat{\beta}_0(\mathbf{t})\}\\
&=\{\mathbb{E}\widehat{\beta}_0(\mathbf{t})-g(\mathbf{t})\}+\{\widetilde{\beta}_0(\mathbf{t})-\mathbb{E}\widetilde{\beta}_0(\mathbf{t})\}+U_n(\mathbf{t})\\
&=\{\mathbb{E}\widehat{\beta}_0(\mathbf{t})-g(\mathbf{t})\}+\frac{A^{1/2}(\mathbf{t},b)}{(nb)^{d/2}}Z_n(\mathbf{t})+U_n(\mathbf{t})\\
&\quad+\bigg(\frac{A^{1/2}(\mathbf{t},b)}{(nb)^{d/2}}-\sqrt{\operatorname{Var}\big(\widetilde{\beta}_0(\mathbf{t})\big)}\bigg)Z_n.
\end{aligned} \tag{49}
$$

The rates of $\mathbb{E}\widehat{\beta}_0(\mathbf{t})-g(\mathbf{t})$ and $U_n(\mathbf{t})$ are provided by Theorem 2 and (48), respectively. By (21) with $x=2$,

$$
\begin{aligned}
\frac{A^{1/2}(\mathbf{t},b)}{(nb)^{d/2}}-\sqrt{\operatorname{Var}\big(\widetilde{\beta}_0(\mathbf{t})\big)}&=(nb)^{-d/2}\Big(A^{1/2}(\mathbf{t},b)-\sqrt{(nb)^d\operatorname{Var}\big(\widetilde{\beta}_0(\mathbf{t})\big)}\Big)\\
&=o\big((nb)^{-d/2}\big),
\end{aligned}
$$

which completes the proof.

4.5. *Proof of Theorem* 5. For $\mathbf{u}=(u_1,\dots,u_d),\mathbf{v}=(v_1,\dots,v_d)\in\mathbb{R}^d$ and $x\in\mathbb{R}$, $\mathbf{u}\le\mathbf{v}$ means $u_j\le v_j$ for all j and $\mathbf{u}\le x$ means $u_j\le x$ for all j. Also, for $\mathbf{u}\le\mathbf{v}$ let $R(\mathbf{u},\mathbf{v})=\{\mathbf{w}:\mathbf{u}\le\mathbf{w}\le\mathbf{v}\}$.

The following result is a byproduct of the proof of Theorem 3.

LEMMA 2. *Assume that* [W1]–[W5] *hold. Then for sufficiently large* n, *there is a constant* $\mathbf{c}_9>0$ *such that*

$$
(nb)^{-dx/2}\bigg|\mathbb{E}\bigg\{\sum_{\mathbf{t}_i\in R(\mathbf{u},\mathbf{v})}\big(\widetilde{W}_n^2(\mathbf{t}_i)-\mathbb{E}\widetilde{W}_n^2(\mathbf{t}_i)\big)\bigg\}^x\bigg|\le(2x-1)!!\mathbf{c}_9^x,
$$

uniformly for all $x\ge2$ *and* $\mathbf{u},\mathbf{v}\in[0,1]^2$ *with* $|u_i-v_i|<2b$ *for* $i\le d$.

PROOF. As in the proof of Theorem 3, write

$$\mathbb{E}\bigg\{\sum_{\mathbf{t}_i \in R(\mathbf{u},\mathbf{v})} \big(\widetilde{W}_n^2(\mathbf{t}_i) - \mathbb{E}\widetilde{W}_n^2(\mathbf{t}_i)\big)\bigg\}^x = \sum_{\mathbf{t}_{i_1},\dots,\mathbf{t}_{i_x} \in R(\mathbf{u},\mathbf{v})} \sum_{\mathbf{S}} \delta_{s_1 s_2}\delta_{s_3 s_4}\cdots\delta_{s_{2x-1}s_{2x}},$$

where $\mathbf{S}$ is the set of all possible ways to make x pairs, $\{(s_1, s_2), \dots, (s_{2x-1}, s_{2x})\}$, with s_j chosen from $\mathbf{I}^x = \{i_1, i_1, i_2, i_2, \dots, i_x, i_x\}$ without replacement and each pair must be of different indices. By (46),

$$(nb)^{-xd/2}\bigg|\sum_{\mathbf{t}_{i_1},\dots,\mathbf{t}_{i_x} \in R(\mathbf{u},\mathbf{v})} \delta_{s_1 s_2}\delta_{s_3 s_4}\cdots\delta_{s_{2x-1}s_{2x}}\bigg| \le \mathbf{c}_8^x (nb)^{-(x-2m)d/2}, \tag{50}$$

where m is the number of chains in $s_1, \dots, s_{2x}$. If x is even, the right-hand side is maximized if $m = x/2$. Taking into account that the number of pairings in $\mathbf{S}$ is bounded above by $(2x-1)!!$, we have for x even,

$$(nb)^{-xd/2}\mathbb{E}\bigg\{\sum_{\mathbf{t}_i \in R(\mathbf{u},\mathbf{v})} \big(\widetilde{W}_n^2(\mathbf{t}_i) - \mathbb{E}\widetilde{W}_n^2(\mathbf{t}_i)\big)\bigg\}^x \le \mathbf{c}_8^x (2x-1)!!. \tag{51}$$

For x odd, by the Cauchy–Schwarz inequality,

$$\begin{aligned}
&\bigg|\mathbb{E}\bigg\{\sum_{\mathbf{t}_i \in R(\mathbf{u},\mathbf{v})} \big(\widetilde{W}_n^2(\mathbf{t}_i) - \mathbb{E}\widetilde{W}_n^2(\mathbf{t}_i)\big)\bigg\}^x\bigg| \\
&\quad \le \mathbb{E}^{1/2}\bigg\{\sum_{\mathbf{t}_i \in R(\mathbf{u},\mathbf{v})} \big(\widetilde{W}_n^2(\mathbf{t}_i) - \mathbb{E}\widetilde{W}_n^2(\mathbf{t}_i)\big)\bigg\}^{x-1} \\
&\qquad \times \mathbb{E}^{1/2}\bigg\{\sum_{\mathbf{t}_i \in R(\mathbf{u},\mathbf{v})} \big(\widetilde{W}_n^2(\mathbf{t}_i) - \mathbb{E}\widetilde{W}_n^2(\mathbf{t}_i)\big)\bigg\}^{x+1},
\end{aligned}$$

and we apply (51) to obtain

$$(nb)^{-xd/2}\bigg|\mathbb{E}\bigg\{\sum_{\mathbf{t}_i \in R(\mathbf{u},\mathbf{v})} \big(\widetilde{W}_n^2(\mathbf{t}_i) - \mathbb{E}\widetilde{W}_n^2(\mathbf{t}_i)\big)\bigg\}^x\bigg| \le \mathbf{c}_8^x \sqrt{2x+1}(2x-1)!!.$$

Since $\sqrt{2x+1} \le 2^x$, the theorem holds if we set $\mathbf{c}_9 = 2\mathbf{c}_8$. □

LEMMA 3. *Assume that* [K′] *and* [W1]–[W5] *hold. Also assume that n and b satisfy $b \to 0$, $nb \to \infty$. Then for some $C < \infty$,*

$$\sup_{\mathbf{t}\in\Omega}\bigg|\frac{1}{(nb)^d}\sum_i k\bigg(\frac{\mathbf{t}_i - \mathbf{t}}{b}\bigg)\big(\widetilde{W}_n^2(\mathbf{t}_i) - \mathbb{E}\widetilde{W}_n^2(\mathbf{t}_i)\big)\bigg| \le C(nb)^{-d/2}\log n \tag{52}$$

and

$$\sup_{\mathbf{t}\in\Omega}\bigg|\frac{1}{(nb)^d}\sum_i k\bigg(\frac{\mathbf{t}_i - \mathbf{t}}{b}\bigg)\widetilde{W}_n(\mathbf{t}_i)\bigg| \le C(nb)^{-d/2}\log n \tag{53}$$

eventually with probability 1 *as $n \to \infty$.*

PROOF. We first consider (52). Let $\check{\beta}_0(\mathbf{t}) = \frac{1}{n^d b^d} \sum_i k(\frac{\mathbf{t}_i - \mathbf{t}}{b}) \widetilde{W}_n^2(\mathbf{t}_i)$. By (K′), $\int_{-1}^1 |k_i'(t)|\, dt < \infty$ for $i = 1, 2, \ldots, d$ and so k_i has bounded variation. Then we can write $k_i = k_{i,1} - k_{i,2}$ where $k_{i,1}$ and $k_{i,2}$ are monotone increasing with $k_{i,1}(-b) = k_{i,2}(-b) = 0$. Define

$$\check{\beta}_{0\mathbf{j}}(\mathbf{t}) = \frac{1}{n^d b^d} \sum_i \prod_{\ell=1}^d k_{\ell, j_\ell}(t_{i\ell} - t_\ell) \widetilde{W}_n^2(\mathbf{t}_i),$$

where $\mathbf{j} = (j_1, \ldots, j_d)$ with $j_\ell \in \{1, 2\}$, $\ell = 1, \ldots, d$. Since $\beta_{0\mathbf{j}}(\mathbf{t})$ is a finite linear combination of $\check{\beta}_{0\mathbf{j}}(\mathbf{t})$, we can just focus on the latter.

Let $\mathbf{1}(\cdot)$ denote the indicator function, and write

$$\begin{aligned}
\check{\beta}_{0\mathbf{j}}(\mathbf{t}) &= \frac{1}{n^d b^d} \sum_i \mathbf{1}(-b \le \mathbf{t}_i - \mathbf{t} \le b) \prod_{\ell=1}^d \int_{-b}^{t_{i\ell} - t_\ell} dk_{\ell, j_\ell}(v_\ell) \widetilde{W}_n^2(\mathbf{t}_i) \\
&= \frac{1}{n^d b^d} \int_{[-b,b]^d} \sum_i \widetilde{W}_n^2(\mathbf{t}_i) \mathbf{1}(\mathbf{v} \le \mathbf{t}_i - \mathbf{t} \le b)\, dk_{1,j_1}(v_1) \cdots dk_{d,j_d}(v_d) \\
&= \frac{1}{n^d b^d} \int_{[-b,b]^d} \sum_i \widetilde{W}_n^2(\mathbf{t}_i) \mathbf{1}(\mathbf{t} + \mathbf{v} \le \mathbf{t}_i \le \mathbf{t} + b)\, dk_{1,j_1}(v_1) \cdots dk_{d,j_d}(v_d).
\end{aligned}$$

Now let

$$G_n(\mathbf{u}, \mathbf{v}) = \frac{1}{n^d b^d} \sum_i \widetilde{W}_n^2(\mathbf{t}_i) \mathbf{1}(\mathbf{u} \le \mathbf{t}_i \le \mathbf{v}) \quad \text{and} \quad G(\mathbf{u}, \mathbf{v}) = \mathbb{E} G_n(\mathbf{u}, \mathbf{v}).$$

Then

$$\begin{aligned}
&\widehat{\beta}_{0\mathbf{j}}(\mathbf{t}) - \mathbb{E}\widehat{\beta}_{0\mathbf{j}}(\mathbf{t}) \\
&\qquad = \int_{[-b,b]^d} \big(G_n(\mathbf{t} + \mathbf{v}, \mathbf{t} + b) - G(\mathbf{t} + \mathbf{v}, \mathbf{t} + b)\big)\, dk_{1,j_1}(v_1) \cdots dk_{d,j_d}(v_d),
\end{aligned}$$

and we have

$$\begin{aligned}
&\sup_{\mathbf{t} \in \Omega} \big|\check{\beta}_{0\mathbf{j}}(\mathbf{t}) - \mathbb{E}\check{\beta}_{0\mathbf{j}}(\mathbf{t})\big| \\
&\qquad \le \sup_{\mathbf{t} \in \Omega} \sup_{0 \le \mathbf{u} \le 2b} \big|G_n(\mathbf{t}, \mathbf{t} + \mathbf{u}) - G(\mathbf{t}, \mathbf{t} + \mathbf{u})\big| \prod_{\ell=1}^d \int_{-b}^b dk_{\ell, j_\ell},
\end{aligned}$$

where $\prod_{\ell=1}^d \int_{-b}^b dk_{\ell,j_\ell} = \prod_{\ell=1}^d \{k_{\ell,j_\ell}(b) - k_{\ell,j_\ell}(-b)\} < \infty$. Thus, it is sufficient to focus on the rate of $\sup_{\mathbf{t} \in \Omega} \sup_{0 \le \mathbf{u} \le 2b} |G_n(\mathbf{t}, \mathbf{t} + \mathbf{u}) - G(\mathbf{t}, \mathbf{t} + \mathbf{u})|$. Now define the cubes $\prod_{j=1}^d [s_j, s_j + [(2b)^{-1}]^{-1}]$ where $s_j \in \{\ell[(2b)^{-1}]^{-1}, \ell = 0, 1, \ldots, [(2b)^{-1}] - 1\}$. For convenience, denote the cubes as $\mathcal{C}_q$. Divide $\mathcal{G}_n$ into subgrids $\mathcal{G}_{n,q}$ where $\mathcal{G}_{n,q}$ contains all the points in $\mathcal{G}_n \cap \mathcal{C}_q$. The number of points in each $\mathcal{G}_{n,q}$ is asymptotically $(2nb)^d$. Denote the grid point in $\mathcal{G}_{n,q}$ that is closest to $\mathbf{0}$ by $\mathbf{s}_q$. It is easy to see that

$$\begin{aligned}&\sup_{\mathbf{t}\in\Omega}\sup_{0\le\mathbf{u}\le 2b}\big|G_n(\mathbf{t},\mathbf{t}+\mathbf{u})-G(\mathbf{t},\mathbf{t}+\mathbf{u})\big|\\ &\qquad\le 2^{2d}\max_q\max_{\mathbf{t}_i\in\mathcal{G}_{n,q}}\big|G_n(\mathbf{s}_q,\mathbf{t}_i)-G(\mathbf{s}_q,\mathbf{t}_i)\big|.\end{aligned}\tag{54}$$

By Lemma 2 and the Markov inequality, for $\mathbf{t}_i\in\mathcal{G}_{n,q}$,

$$\begin{aligned}&\mathbb{P}\biggl(\frac{2^{2d}(nb)^{d/2}}{\log n}\big|G_n(\mathbf{s}_q,\mathbf{t}_i)-G(\mathbf{s}_q,\mathbf{t}_i)\big|>(d+2)\mathbf{c}_{10}\biggr)\\ &\quad=\mathbb{P}\biggl(\exp\biggl\{\frac{2^{2d}(nb)^{d/2}}{\mathbf{c}_{10}}\big|G_n(\mathbf{s}_q,\mathbf{t}_i)-G(\mathbf{s}_q,\mathbf{t}_i)\big|\biggr\}>e^{(d+2)\log n}\biggr)\\ &\quad\le e^{-(d+2)\log n}\sum_{x=0}^{\infty}\frac{1}{x!}\mathbb{E}\biggl(\frac{2^{2d}(nb)^{d/2}}{\mathbf{c}_{10}}\big|G_n(\mathbf{s}_q,\mathbf{t}_i)-G(\mathbf{s}_q,\mathbf{t}_i)\big|\biggr)^x\\ &\quad\le e^{-(d+2)\log n}\sum_{x=0}^{\infty}\frac{1}{x!}\biggl(\frac{2^{2d}}{\mathbf{c}_{10}}\biggr)^x(2x-1)!!\mathbf{c}_9^x\\ &\quad\le\bigl(1-2^{2d+1}\mathbf{c}_9/\mathbf{c}_{10}\bigr)^{-1}n^{-(d+2)}\end{aligned}$$

so long as $\mathbf{c}_{10}>2^{2d+1}\mathbf{c}_9$. Consequently,

$$\begin{aligned}&\sum_{n=1}^{\infty}\mathbb{P}\biggl(\frac{(nb)^{d/2}}{\log n}\sup_{\mathbf{t}\in\Omega}\sup_{0\le\mathbf{u}\le 2b}\big|G_n(\mathbf{t},\mathbf{t}+\mathbf{u})-G(\mathbf{t},\mathbf{t}+\mathbf{u})\big|>(d+2)\mathbf{c}_{10}\biggr)\\ &\quad\le\bigl(1-2^{2d+1}\mathbf{c}_9/\mathbf{c}_{10}\bigr)^{-1}\sum_{n=1}^{\infty}\sum_q\sum_{\mathbf{t}_i\in\mathcal{G}_{n,q}}n^{-(d+2)}\\ &\quad=\bigl(1-2^{2d+1}\mathbf{c}_9/\mathbf{c}_{10}\bigr)^{-1}\sum_{n=1}^{\infty}n^{-2}<\infty.\end{aligned}$$

Applying the Borel–Cantelli lemma, we conclude that for some $C\in(0,\infty)$,

$$\frac{(nb)^{d/2}}{\log n}\sup_{\mathbf{t}\in\Omega}\sup_{\mathbf{u}\le 2b,\mathbf{u}>0}\big|G_n(\mathbf{t},\mathbf{t}+\mathbf{u})-G(\mathbf{t},\mathbf{t}+\mathbf{u})\big|\le C$$

eventually w.p.1. This completes the proof of (52).

The proof of (53) is similar but easier. Since $\widetilde{W}_n(\mathbf{t})$ is a Gaussian process with mean zero and covariance C_n, for x even we have

$$\mathbb{E}\biggl\{\sum_i k\biggl(\frac{\mathbf{t}_i-\mathbf{t}}{b}\biggr)\widetilde{W}_n(\mathbf{t}_i)\biggr\}^x=x!!\biggl\{\sum_{i,j}k\biggl(\frac{\mathbf{t}_i-\mathbf{t}}{b}\biggr)\biggl(\frac{\mathbf{t}_i-\mathbf{t}}{b}\biggr)C_n(\mathbf{t}_i,\mathbf{t}_j)\biggr\}^{x/2}.$$

Since

$$\sum_{i,j} k\left(\frac{\mathbf{t}_i-\mathbf{t}}{b}\right)\left(\frac{\mathbf{t}_i-\mathbf{t}}{b}\right)C_n(\mathbf{t}_i,\mathbf{t}_j)=O\big((nb)^d\big),$$

there exists some $C<\infty$ such that

$$\mathbb{E}\left\{\frac{1}{(nb)^{d/2}}\sum_i k\left(\frac{\mathbf{t}_i-\mathbf{t}}{b}\right)\widetilde{W}_n(\mathbf{t}_i)\right\}^x \le x!!C^x.$$

The rest of the proof follows in the same way as in (52). $\square$

PROOF OF THEOREM 5. As in (49), write

$$\widehat{\beta}_0(\mathbf{t})-g(\mathbf{t})=\{\mathbb{E}\widehat{\beta}_0(\mathbf{t})-g(\mathbf{t})\}+\{\widetilde{\beta}_0(\mathbf{t})-\mathbb{E}\widetilde{\beta}_0(\mathbf{t})\}+U_n(\mathbf{t}).$$

The first term on the right is the bias, which is handled by Theorem 2. Replacing k with $\breve{k}$ in (52) of Lemma 3, we conclude that there exists $C<\infty$ such that

$$\sup_{\mathbf{t}\in\Omega}\big|\widetilde{\beta}_0(\mathbf{t})-\mathbb{E}\widetilde{\beta}_0(\mathbf{t})\big|\le C(nb)^{-d/2}\log n$$

eventually with probability 1 as $n\to\infty$. Similarly, replacing $k((\mathbf{t}_i-\mathbf{t})/b)$ with $k((\mathbf{t}_i-\mathbf{t})/b)\mu_n(\mathbf{t}_i)$ in (53), we conclude that there exists $C<\infty$ such that

$$\sup_{\mathbf{t}\in\Omega}\big|U_n(\mathbf{t})\big|\le C\delta_n(nb)^{-d/2}\log n$$

eventually with probability 1 as $n\to\infty$. $\square$

4.6. *Deformation process and multifractional Brownian motion*. This subsection is devoted to the proofs of the conditions [R1] and [R2] for the deformation process and multi-fractional Brownian introduced in Section 1.

Deformation process. We first consider the deformation process $Y(\mathbf{t})=X(F(\mathbf{t}))$ where X is stationary and whose covariance of X satisfies the right-hand side of (6) and F is a one-to-one and $2(k+1)$-times differentiable function from $\mathbb{R}^d$ to $\mathbb{R}^d$. Recall that we let $k<\nu<k+1$ for some nonnegative integer k so that $\lceil\nu\rceil=k+1$.

Write $F(\mathbf{t})=(F_1(\mathbf{t}),\dots,F_d(\mathbf{t}))$, and define $F_j^{(\boldsymbol{\ell})}(\mathbf{t})=\frac{\partial^{\ell_d}}{\partial t_d^{\ell_d}}\cdots\frac{\partial^{\ell_1}}{\partial t_1^{\ell_1}}F_j(\mathbf{t})$ where $\frac{\partial^0}{\partial t_j^0}$ is interpreted as 1. Also, for $\mathbf{h}=(h_1,\dots,h_d)\in\mathbb{R}^d$ and $\boldsymbol{\ell}=(\ell_1,\dots,\ell_d)\in\{0,1,2,\dots\}^d$, let $|\boldsymbol{\ell}|=\sum_j\ell_j$ and $\mathbf{h}^{\boldsymbol{\ell}}=h_1^{\ell_1}\cdots h_d^{\ell_d}$. It follows from Taylor's theorem that

$$7F_j(\mathbf{t}+\mathbf{h})-F_j(\mathbf{t})=\sum_{|\boldsymbol{\ell}|=1}^{2k+2}\frac{1}{|\boldsymbol{\ell}|!}F_j^{(\boldsymbol{\ell})}(\mathbf{t})\mathbf{h}^{\boldsymbol{\ell}}+o\big(|\mathbf{h}|^{2k+2}\big).$$

Thus,

$$\big|F(\mathbf{t}+\mathbf{h})-F(\mathbf{t})\big|^2=\sum_{j=1}^{d}\left(\sum_{|\boldsymbol{\ell}|=1}^{2k+2}\frac{1}{|\boldsymbol{\ell}|!}F_j^{(\boldsymbol{\ell})}(\mathbf{t})\mathbf{h}^{\boldsymbol{\ell}}\right)^2+o\big(|\mathbf{h}|^{2k+3}\big),$$

and for $s=1,\dots,k+1$,

$$\begin{aligned}&\big|F(\mathbf{t}+\mathbf{h})-F(\mathbf{t})\big|^{2s}\\&\quad=\sum_{|\boldsymbol{\ell}_1|+\cdots+|\boldsymbol{\ell}_{2s}|=2s}^{2k+2}\frac{\mathbf{h}^{\boldsymbol{\ell}_1+\cdots+\boldsymbol{\ell}_{2s}}}{|\boldsymbol{\ell}_1|!\cdots|\boldsymbol{\ell}_{2s}|!}\prod_{r=1}^{s}\sum_{j=1}^{d}F_j^{(\boldsymbol{\ell}_{2r-1})}(\mathbf{t})F_j^{(\boldsymbol{\ell}_{2r})}(\mathbf{t})+O\big(|\mathbf{h}|^{2k+3}\big).\end{aligned}$$

On the other hand, with $J_F^{\mathbf{t}}$ denoting the Jacobian of F at $\mathbf{t}$,

$$F(\mathbf{t}+\mathbf{h})-F(\mathbf{t})=J_F^{\mathbf{t}}\mathbf{h}+O\big(|\mathbf{h}|^2\big),$$

and hence

$$\big|F(\mathbf{t}+\mathbf{h})-F(\mathbf{t})\big|^{2\nu}=\big|J_F^{\mathbf{t}}\mathbf{h}\big|^{2\nu}+O\big(|\mathbf{h}|^{2\nu+1}\big),$$

where $|J_F^{\mathbf{t}}\mathbf{h}|^2=\sum_{j=1}^{d}(\sum_{|\boldsymbol{\ell}|=1}F_j^{(\boldsymbol{\ell})}(\mathbf{t})\mathbf{h}^{\boldsymbol{\ell}})^2$. Since $2\nu+1\le\lceil 2\nu+1\rceil=\lceil 2\nu\rceil+1$, we can write

$$C(\mathbf{t},\mathbf{t}+\mathbf{h})=\begin{cases}c_0+c_{2\nu}\big|J_F^{\mathbf{t}}\mathbf{h}\big|^{2\nu}+O\big(|\mathbf{h}|^{2\nu+1}\big), & \nu\le 1/2,\\ c_0+\displaystyle\sum_{|\boldsymbol{\ell}|=2}^{\lceil 2\nu\rceil}b_{\boldsymbol{\ell}}(\mathbf{t})\mathbf{h}^{\boldsymbol{\ell}}+c_{2\nu}\big|J_F^{\mathbf{t}}\mathbf{h}\big|^{2\nu}+O\big(|\mathbf{h}|^{2\nu+1}\big) & \nu>1/2,\end{cases}\tag{55}$$

as $\mathbf{h}\to\mathbf{0}$ for some functions $b_{\boldsymbol{\ell}}$. Using the fact that the covariance function is symmetric, it is straightforward to verify that [R1] holds with $r=\lceil 2\nu\rceil,\alpha(\mathbf{t})\equiv 2\nu,\gamma(\mathbf{t})\equiv 1$ and $\Psi_{\mathbf{t}}(\mathbf{h})=c_{2\nu}|J_F^{\mathbf{t}}\mathbf{h}|$.

Multifractional Brownian motion. Assume that $H(\mathbf{t})$ is three times differentiable and $|\mathbf{t}|$ bounded away from 0. It is easy to verify that

$$D^{(k)}(x)=\int_{\mathbb{R}^d}\frac{(1-e^{iu_1})(-\log|\mathbf{u}|)^k}{|\mathbf{u}|^{x+d}}\,d\mathbf{u},$$

which is well defined for $x\in(0,1)$. Holding $\mathbf{t}$ fixed, it follows that

$$\begin{aligned}|\mathbf{t}|^{H(\mathbf{s})+H(\mathbf{t})}&=|\mathbf{t}|^{2H(\mathbf{t})}e^{\{H(\mathbf{s})-H(\mathbf{t})\}\log|\mathbf{t}|}\\&=|\mathbf{t}|^{2H(\mathbf{t})}\left(1+\sum_{|\boldsymbol{\ell}|=1}^{2}\frac{f^{(\boldsymbol{\ell})}(\mathbf{t})(\mathbf{s}-\mathbf{t})^{\boldsymbol{\ell}}}{|\boldsymbol{\ell}|!}+O\big(|\mathbf{s}-\mathbf{t}|^3\big)\right)\end{aligned}$$

and

$$D\big(H(\mathbf{s})+H(\mathbf{t})\big)=\left(D\big(2H(\mathbf{t})\big)+\sum_{|\boldsymbol{\ell}|=1}^{2}\frac{g^{(\boldsymbol{\ell})}(\mathbf{t})(\mathbf{s}-\mathbf{t})^{\boldsymbol{\ell}}}{|\boldsymbol{\ell}|!}+O\big(|\mathbf{s}-\mathbf{t}|^3\big)\right),$$

where $f(\mathbf{s}) = e^{\{H(\mathbf{s})-H(\mathbf{t})\}\log|\mathbf{t}|}$ and $g(\mathbf{s}) = D(H(\mathbf{s}) + H(\mathbf{t}))$. Thus,

$$D\big(H(\mathbf{s}) + H(\mathbf{t})\big)|\mathbf{t}|^{H(\mathbf{s})+H(\mathbf{t})} = \sum_{|\boldsymbol{\ell}|=0}^{2} b_{\boldsymbol{\ell}}(\mathbf{t})\mathbf{s}^{\boldsymbol{\ell}} + \big(|\mathbf{t}|^{2H(\mathbf{t})} + D\big(2H(\mathbf{t})\big)\big)O\big(|\mathbf{s}-\mathbf{t}|^3\big),$$

and, by symmetry, we also have

$$D\big(H(\mathbf{s}) + H(\mathbf{t})\big)|\mathbf{s}|^{H(\mathbf{s})+H(\mathbf{t})} = \sum_{|\boldsymbol{\ell}|=0}^{2} b_{\boldsymbol{\ell}}(\mathbf{s})\mathbf{t}^{\boldsymbol{\ell}} + \big(|\mathbf{s}|^{2H(\mathbf{s})} + D\big(2H(\mathbf{s})\big)\big)O\big(|\mathbf{s}-\mathbf{t}|^3\big).$$

Similarly,

$$\begin{aligned} &D\big(H(\mathbf{s}) + H(\mathbf{t})\big)|\mathbf{s}-\mathbf{t}|^{H(\mathbf{s})+H(\mathbf{t})} \\ &\quad = |\mathbf{s}-\mathbf{t}|^{2H(\mathbf{t})}D\big(2H(\mathbf{t})\big) + O\big(|\mathbf{s}-\mathbf{t}|^{2H(\mathbf{t})+1}\log\big(|\mathbf{s}-\mathbf{t}|\big)\big). \end{aligned}$$

It is straightforward to check that $r = 2$, $\Psi_{\mathbf{t}}(\mathbf{h}) = D(2H(\mathbf{t}))$ for all $\mathbf{h}$, $\alpha(\mathbf{t}) = 2H(\mathbf{t})$ and γ is any constant less than 1.

Acknowledgments. We are very grateful to the referees for their comments and suggestions.

REFERENCES

ANDERES, E. and CHATTERJEE, S. (2009). Consistent estimates of deformed isotropic Gaussian random fields on the plane. *Ann. Statist.* **37** 2324–2350. MR2543694

ANDERES, E. B. and STEIN, M. L. (2008). Estimating deformations of isotropic Gaussian random fields on the plane. *Ann. Statist.* **36** 719–741. MR2396813

AYACHE, A., SHIEH, N.-R. and XIAO, Y. (2011). Multiparameter multifractional Brownian motion: Local nondeterminism and joint continuity of the local times. *Ann. Inst. Henri Poincaré Probab. Stat.* **47** 1029–1054. MR2884223

CHAN, G. and WOOD, A. T. A. (2000). Increment-based estimators of fractal dimension for two-dimensional surface data. *Statist. Sinica* **10** 343–376. MR1769748

CHILÈS, J.-P. and DELFINER, P. (2012). *Geostatistics: Modeling Spatial Uncertainty*, 2nd ed. Wiley, Hoboken, NJ. MR2850475

CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*. Wiley, New York. MR1239641

DAVIES, S. and HALL, P. (1999). Fractal analysis of surface roughness by using spatial data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61** 3–37. MR1664088

DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory: An Introduction*. Springer, New York. MR2234156

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications. Monographs on Statistics and Applied Probability* **66**. Chapman & Hall, London. MR1383587

FUENTES, M. (2001). A new high frequency kriging approach for nonstationary environmental processes. *Envirometrics* **12** 469–483.

FUENTES, M. (2002). Spectral methods for nonstationary spatial processes. *Biometrika* **89** 197–210. MR1888368

FUGLSTAD, G.-A., LINDGREN, F., SIMPSON, D. and RUE, H. (2013). Non-stationary spatial modeling with applications to spatial prediction and precipitation. Manuscript.

FUGLSTAD, G.-A., LINDGREN, F., SIMPSON, D. and RUE, H. (2015). Exploring a new class of non-stationary spatial Gaussian random fields with varying local anisotropy. *Statist*. *Sinica* **25** 115–133. MR3328806

HERBIN, E. (2006). From N parameter fractional Brownian motions to N parameter multifractional Brownian motions. *Rocky Mountain J*. *Math*. **36** 1249–1284. MR2274895

HIGDON, D., SWALL, J. and KERN, J. (1999). Nonstationary spatial modeling. *Bayesian Stat*. **6** 761–768.

ISSERLIS, L. (1918). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika* **12** 134–139.

KENT, J. T. and WOOD, A. T. A. (1997). Estimating the fractal dimension of a locally self-similar Gaussian process by using increments. *J*. *Roy*. *Statist*. *Soc*. *Ser*. *B* **59** 679–699. MR1452033

KIM, H.-M., MALLICK, B. K. and HOLMES, C. C. (2005). Analyzing nonstationary spatial data using piecewise Gaussian processes. *J*. *Amer*. *Statist*. *Assoc*. **100** 653–668. MR2160567

MANDELBROT, B. B. and VAN NESS, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev*. **10** 422–437. MR0242239

MATHERON, G. (1964). Equation de la chaleur, écoulements en milieu poreux et diffusion géochimique. Internal Report (Note Géostatistique 55), BRGM.

MATHERON, G. (1973). The intrinsic random functions and their applications. *Adv*. *in Appl*. *Probab*. **5** 439–468. MR0356209

SAMPSON, P. D. (2010). Constructions for nonstationary spatial processes. In *Handbook of Spatial Statistics* (A. E. Gelfand, P. J. Diggle, M. Fuentes and P. Guttorp, eds.) 119–130. CRC Press, Boca Raton, FL. MR2730945

SAMPSON, P. D. and GUTTORP, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *J*. *Amer*. *Statist*. *Assoc*. **87** 108–119.

STEIN, M. L. (1999). *Interpolation of Spatial Data*: *Some Theory for Kriging*. Springer, New York. MR1697409

STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann*. *Statist*. **10** 1040–1053. MR0673642

WHITTLE, P. (1954). On stationary processes in the plane. *Biometrika* **41** 434–449. MR0067450

T. HSING
B. THELEN
DEPARTMENT OF STATISTICS
1085 SOUTH UNIVERSITY
ANN ARBOR, MICHIGAN 48109-1107
USA
E-MAIL: thsing@umich.edu
bjthelen@umich.edu

T. BROWN
39100 COUNTRY CLUB DRIVE
FARMINGTON HILLS, MICHIGAN 48331
USA
E-MAIL: tommybro@umich.edu