Analzing Spatial Processes Locally

by

Thomas Anthony Brown

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Statistics)
in The University of Michigan
2014

Doctoral Committee:

Professor Tailen Hsing, Chair
Assistant Professor Veronica Berrocal
Associate Professor Stilian Stoev
Lecturer Brian Thelen

For my wife Kandra, my brother Travis and my parents Tom and Regina

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF APPENDICES

Appendix

LIST OF SYMBOLS

$\mathbb{R}$	The real numbers.
$\mathbb{Z}$	The integers. $\mathbb{N}$ denotes the non-negative integers.
$\boldsymbol{t}$	Spatial location in $\mathbb{R}^d$, $\boldsymbol{t} = (t_1, t_2, \ldots, t_d)$. Note that $\boldsymbol{s}, \boldsymbol{z}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{u}, \boldsymbol{h}, \boldsymbol{v}$ are defined similarly.
$\boldsymbol{t}_i$	Spatial location for index i, $\boldsymbol{t}_i = (t_{i1}, t_{i2}, \ldots, t_{id})$.
$\boldsymbol{\ell}$	Vector in $\mathbb{N}^d$, $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_d)$. Note that $\boldsymbol{t}^{\boldsymbol{\ell}} = (t_1^{\ell_1}, \ldots, t_d^{\ell_d})$ and $\boldsymbol{t}^{\ell} = (t_1^{\ell}, \ldots, t_d^{\ell})$.
$\lvert \cdot \rvert$	Eulidean distance operator. Note that for $\boldsymbol{\ell}$ we define this as $\lvert\boldsymbol{\ell}\rvert = \sum_{i=1}^{d} \ell_i$.
$\det(\cdot)$	Determinant function.
$\lfloor \ell \rfloor$	Greatest integer lower than ℓ. Similarly $\lceil \ell \rceil$ is the smallest integer greater than ℓ.
$k!$	k factorial, equal to $k(k-1)\cdots 2 \cdot 1$.
$k!!$	Double factorial, equal to $k(k-2)\cdots 4 \cdot 2$ for k even and $k(k-2)\cdots 3 \cdot 1$ for k odd.
$\binom{n}{k}$	n "choose" k, equal to $\left(\frac{n!}{k!(n-k)!}\right)$.
$\frac{\partial}{\partial t_i} g(\boldsymbol{t})$	Partial derivative w.r.t. dimension i of $\boldsymbol{t}$.
$\partial_{\boldsymbol{u}} g(\boldsymbol{t})$	Directional derivative of g in the direction $\boldsymbol{u}$ at $\boldsymbol{t}$.
$\partial_{\boldsymbol{u}}^{(i,j)} g(\boldsymbol{t}, \boldsymbol{s})$	The partial derivative of g in the direction $\boldsymbol{u}$ w.r.t $\boldsymbol{t}$ i times and w.r.t. $\boldsymbol{s}$ j times.
$\nabla g(\boldsymbol{t})$	The gradient of g at $\boldsymbol{t}$.

$H_g^{\boldsymbol{t}}$	The Hessian of g at $\boldsymbol{t}$.
$J_g^{\boldsymbol{t}}$	The Jacobian of g at $\boldsymbol{t}$.
$O(\cdot)$	Big-O notation. If $\lvert F(\boldsymbol{t})/G(\boldsymbol{t})\rvert \leq c$ for some finite $c > 0$ and t sufficiently large or small then write $F(t) = O(G(t))$. If the inequality holds for all $c > 0$, then $F(\boldsymbol{t}) = o(G(\boldsymbol{t}))$.
$O_p(\cdot)$	Convergence in probability.
$\Gamma(\cdot)$	The Gamma function.
$\overset{d}{=}$	Equal in distribution.
$\rightarrow$	"Converges to".
$\Rightarrow$	"Implies".
$\approx$	"Approximately".
$\asymp$	$F(\boldsymbol{t}) \asymp G(\boldsymbol{t})$ if $F(\boldsymbol{t}) = O(G(\boldsymbol{t}))$ and $G(\boldsymbol{t}) = O(F(\boldsymbol{t}))$.
$\Delta_{\boldsymbol{h}}^x$	Increment operator of order x in the direction $\boldsymbol{h}$. $\Delta_{\boldsymbol{h}} Y(\boldsymbol{t}) = Y(\boldsymbol{t}+\boldsymbol{h}) - Y(\boldsymbol{t})$ if $x = 1$ and $\Delta_{\boldsymbol{h}}^x Y(\boldsymbol{t}) = \Delta_{\boldsymbol{h}}^{x-1}\big(Y(\boldsymbol{t}+\boldsymbol{h}) - Y(\boldsymbol{t})\big)$ when $x \geq 2$.
$\sup_{\boldsymbol{t}\in\Omega}$	The supremum over $\boldsymbol{t}$ in Ω.
$k(\cdot)$	A kernel function.

ABSTRACT

Analyzing Spatial Processes Locally
by
Thomas Brown

Chair: Tailen Hsing

The emergence of dense spatial data sets allows us to examine spatial processes on a local level. This thesis analyzes local prediction and local estimation of the covariance model for a Gaussian process observed on a single dense regular grid. We assume a smooth mean and make the assumption that locally, the covariance function is stationary and approximately an even polynomial plus a principal irregular term. This covariance model is applicable to a large class of processes including some which are locally stationary, but nonstationary globally, and processes with locally stationary increments. Some examples include deformation models with stationary autocovariances (e.g., Matérn) and multifractional Brownian motion.

We justify the use of a Kriging estimator which relies on the covariance only through the principal term. Then we consider local estimation of the principal term through a local linear smoother and prove infill asymptotic convergence results. We prove a central limit theorem with a rate matching the optimal nonparametric rate assuming two derivatives and prove an almost sure uniform convergence result with a rate slightly slower than optimal. Simulation results are provided that validate our theory and we explore additional problems such as estimation at the boundary and missing data.

CHAPTER 1

Introduction

In the world of spatial data, data comprised of measurements of a statistical process with a corresponding spatial location, the most common analyses considered are prediction at unobserved spatial locations and estimation of parameters which govern the process. Among prediction techniques, a popular choice is Kriging, the best linear unbiased predictor. For Gaussian processes, Kriging relies on the mean and covariance structure of the spatial process under consideration. If the mean or covariance structure is unknown, then they need to be assumed or estimated from the data. In this thesis, we make some general local stationarity assumptions on an observed spatial process, develop a corresponding Kriging estimator and prove consistency results for estimates of the relevant covariance parameters needed for Kriging.

We assume that the spatial process of interest Y belongs to the class of Gaussian processes where the mean μ smoothly changes with spatial location and the covariance between two points $C(\boldsymbol{t}, \boldsymbol{s})$, $\boldsymbol{t}, \boldsymbol{s} \in \mathbb{R}^d$ satisfies the property

$$C(\boldsymbol{t}, \boldsymbol{s}) = p(\boldsymbol{t}, \boldsymbol{s}) + f_{\boldsymbol{t}}(\boldsymbol{t}-\boldsymbol{s})|\boldsymbol{t}-\boldsymbol{s}|^{\alpha} + O\left(|\boldsymbol{t}-\boldsymbol{s}|^{\alpha+\gamma}\right), \tag{1.1}$$

where $O(|\boldsymbol{t}-\boldsymbol{s}|)$ is as $|\boldsymbol{t}-\boldsymbol{s}| \to 0$, α is a smoothness parameter(possibly varying with $\boldsymbol{t}$), $\alpha/2 > 0$ is non-integer, $f_{\boldsymbol{t}}(\boldsymbol{t}-\boldsymbol{s})$ is a smooth function that depends on $\boldsymbol{t}-\boldsymbol{s}$ only through its direction, $p(\boldsymbol{t}, \boldsymbol{s})$ is a polynomial in $\boldsymbol{t}$ and $\boldsymbol{s}$ and $\gamma > 0$. The parameter α is a measure of smoothness of Y, where sample paths of Y are $\lfloor \alpha/2 \rfloor$ times mean square differentiable. $\alpha/2$ is also referred to as the Hurst parameter and is a measure of long-term dependence in time series data or surface roughness in higher dimensional dimensional random processes. $f_{\boldsymbol{t}}(\boldsymbol{t}-\boldsymbol{s})|\boldsymbol{t}-\boldsymbol{s}|^{\alpha}$ is referred to as the principal irregular term.

This class of processes includes anisotropic Matérn Gaussian random fields (Anderes 2010), multifractional Brownian motion (Ayache 2000, Herbin 2006), the deformation model (Sampson and Guttorp 1992, Anderes and Stein 2008, Anderes and Chaterjee 2009) and intrinsic random functions (Matheron 1964, Chilés and Delfiner 1999).

By taking high order increments of Y, we remove the low order behavior of μ and the polynomial function $p(\boldsymbol{t}, \boldsymbol{s})$ in C. It follows that squared increments will approximate the principal irregular term. The use of squared increments began with the quadratic variation theorem in Levy (1940). He showed that for Brownian motion on $[0, 1]$, the average of the squared increments of order 1 converges a.s. to 1. Later, Baxter (1956) showed a.s. convergence for stochastic processes on $[0, 1]$ with Gaussian increments whose mean function has bounded first derivative and covariance function has bounded second derivative. Kozin (1957) extended this result to a limit theorem when the processes increments are stationary and independent and Gladyshev(1961) extended it to a larger class of processes with Gaussian increments. Istas and Lang (1997) use squared increments to estimate the local Hölder index when $d = 1$. Anderes (2010) uses them to estimate the parameters of a Gaussian random field with a geometric anisotropic Matérn autocovariance, and show that the scale and variance parameters can be consistently estimated separately when $d > 4$.

Rosenblatt (1956) introduced kernel estimators for density estimation. Nadaraya (1964) and Watson (1964) used kernel estimators for regression estimation. Local linear estimators have been used by Stone (1977) and Fan et. al. (1997). We evaluate the local linear kernel regression estimator of the principal irregular term, which is a local linear smoothing of the squared increments of Y within some window. We prove infill asymptotic convergence rates for this estimator, that is, we prove convergence as the width of the smoothing window decreases to 0 and the number of points in the window simultaneously increases to ∞.

Anderes and Chatterjee (2009) use a kernel smoother to estimate the deformation of an isotropic random field with the restriction that $\alpha < 2$. We show that this restriction can be extended to non-even $\alpha > 2$ if we assume the deformation is sufficiently smooth. We also show that the local linear kernel smoother has a smaller bias than the Nadaraya-Watson estimator when estimating close to the boundary.

Stone (1982) establishes that $n^{(p-m)/(2p+d)}$ is the optimal nonparametric convergence rate for estimating the m^{th} derivative of an unknown regresion function assuming i.i.d errors, p derivatives, dimension d and sample size n. He also shows that the optimal a.s. convergence rate is $(n^{-1}\log(n))^{(p-m)/(2p+d)}$. We show that the local linear kernel regression estimator matches Stone's standard nonparametric rate assuming two derivatives and has a slightly smaller than optimal a.s. convergence rate.

The rest of the thesis is structured as follows. Later in this chapter, we give an introduction to the relevant topics in statistics that we utilize in this thesis. In chapter 2, we present examples of processes which belong to this class of processes and justify the use of a Kriging estimator similar to universal Kriging and intrinsic random function Kriging. In chapter 3, we prove consistency results for the local linear kernel regression estimator. Chapter 4 explores related topics such as estimation at the boundary, missing data and estimation of α. Chapter 5 provides simulation results to confirm our theoretical calculations. The appendices provide the proofs from the earlier chapters.

1.1 Gaussian Processes

Suppose you have measurements of a spatial process $Y(\boldsymbol{t}) \in \mathbb{R}$ at locations $\boldsymbol{t}_i$, $i \in \{1, 2, \ldots, N\}$, where $\boldsymbol{t}, \boldsymbol{t}_i \in \Omega$, for some spatial domain $\Omega \subset \mathbb{R}^d$, $d \in \{1, 2, \ldots\}$. Then for example, you may wish to predict $Y(\boldsymbol{t}_0)$ at some unobserved location $\boldsymbol{t}_0 \in \Omega$ or estimate some parameter which specifies the covariance function of the process.

In statistics, $Y(\boldsymbol{t})$ is typically assumed to be a random function or a random variable which is characterized by a distribution function. Let

$$P_{\boldsymbol{t}_1, \boldsymbol{t}_2, \ldots, \boldsymbol{t}_N}(z_1, z_2, \ldots, z_N) := P(Y(\boldsymbol{t}_1) \leq z_1, Y(\boldsymbol{t}_2) \leq z_2, \ldots, Y(\boldsymbol{t}_N) \leq z_N)$$

denote the $N-$variate cdf of Y for spatial locations $\boldsymbol{t}_i \in [0,1]^d$ and numbers $z_i \in \mathbb{R}$ for $i = 1, \ldots, N$. Letting $p_{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_N}$ denote the corresponding $N-$variate pdf, define the expected value of $Y(\boldsymbol{t})$ as

$$\mu(\boldsymbol{t}) := E(Y(\boldsymbol{t})) = \int_{-\infty}^{\infty} z p_{\boldsymbol{t}}(z) dz$$

and define the covariance of Y between two spatial locations $\boldsymbol{t}$ and $\boldsymbol{s}$ as

$$\begin{aligned} C(\boldsymbol{t}, \boldsymbol{s}) &:= E\left[(Y(\boldsymbol{t}) - \mu(\boldsymbol{t}))(Y(\boldsymbol{s}) - \mu(\boldsymbol{s}))\right] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (z - \mu(\boldsymbol{t}))(x - \mu(\boldsymbol{s})) p_{\boldsymbol{t}, \boldsymbol{s}}(z, x) dx, dz. \end{aligned}$$

From this we can define the variance of Y at $\boldsymbol{t}$ as $\mathrm{Var}(Y(\boldsymbol{t})) := C(\boldsymbol{t}, \boldsymbol{t})$.

It is common in spatial statistics to assume that the observed process is Gaussian. For Gaussian processs,

$$p_{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_N}(y_1, \ldots, y_N) = (2\pi)^{-N/2} \det(\Sigma)^{-1/2} e^{-\frac{1}{2}(\boldsymbol{y} - \mu)' \Sigma^{-1} (\boldsymbol{y} - \mu)}$$

where $\boldsymbol{y} = (y_1, \ldots, y_N)'$, $\mu = (\mu(\boldsymbol{t}_1), \ldots, \mu(\boldsymbol{t}_N))'$ and

$$\Sigma = \begin{pmatrix} C(\boldsymbol{t}_1, \boldsymbol{t}_1) & C(\boldsymbol{t}_1, \boldsymbol{t}_2) & \cdots & C(\boldsymbol{t}_1, \boldsymbol{t}_N) \\ C(\boldsymbol{t}_2, \boldsymbol{t}_1) & C(\boldsymbol{t}_2, \boldsymbol{t}_2) & \cdots & C(\boldsymbol{t}_2, \boldsymbol{t}_N) \\ \vdots & \vdots & \ddots & \vdots \\ C(\boldsymbol{t}_N, \boldsymbol{t}_1) & C(\boldsymbol{t}_N, \boldsymbol{t}_2) & \cdots & C(\boldsymbol{t}_N, \boldsymbol{t}_N) \end{pmatrix}.$$

We require that Σ is positive definite so that $\det(\Sigma) > 0$.

One benefit of the Gaussian assumption is that Y is fully characterized by μ and C. Further, we can write

$$Y(\boldsymbol{t}) \stackrel{d}{=} \mu(\boldsymbol{t}) + Z(\boldsymbol{t})$$

where $Z(\boldsymbol{t})$ is a mean 0 Gaussian process with covariance function C.

Another desirable consequence of the Gaussian assumption is that we can directly calculate bias and variance for estimators of Y and develop consistency results for estimators of covariance parameters. As Cressie (1993) points out, the Gaussian assumption makes analysis much more simple and the average of many small and possibly non-Gaussian effects will be approximately Gaussian by the central limit theorem.

1.2 Differentiability

Let $\Delta_{\boldsymbol{u}} g(\boldsymbol{t}) := g(\boldsymbol{t} + \boldsymbol{u}) - g(\boldsymbol{t})$ denote the increment in the direction $\boldsymbol{u}$. Then let $\Delta^x_{\boldsymbol{u}} g(\boldsymbol{t}) := \Delta_{\boldsymbol{u}} \Delta^{x-1}_{\boldsymbol{h}} g(\boldsymbol{t})$ denote the increment of order x in the direction $\boldsymbol{u}$. The directional derivative of g in the direction $\boldsymbol{u}$ at the point $\boldsymbol{t}$ is defined as

$$\begin{aligned} \lim_{n\to\infty} n\Delta_{\boldsymbol{u}/n} g(\boldsymbol{t}) &= \sum_{i=1}^{d} u_i \frac{\partial}{\partial t_i} g(\boldsymbol{t}) \\ &:= \partial_{\boldsymbol{u}} g(\boldsymbol{t}). \end{aligned}$$

The k^{th} order mixed parial derivative is defined as $\frac{\partial}{\partial t_{i_1}} \cdots \frac{\partial}{\partial t_{i_k}} g(\boldsymbol{t})$ where $i_j \in \{1, \ldots, d\}$, $j \in \{1, \ldots, k\}$. Then the k^{th} order directional derivative in directions $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$ is defined as

$$\begin{aligned} \lim_{n\to\infty} n^k \Delta_{\boldsymbol{u}_1/n} \cdots \Delta_{\boldsymbol{u}_k/n} g(\boldsymbol{t}) &= \sum_{i_1=1}^{d} \cdots \sum_{i_k=1}^{d} u_{1i_1} \cdots u_{ki_k} \frac{\partial}{\partial t_{i_1}} \cdots \frac{\partial}{\partial t_{i_k}} g(\boldsymbol{t}) \\ &:= \partial_{\boldsymbol{u}_1} \cdots \partial_{\boldsymbol{u}_k} g(\boldsymbol{t}). \end{aligned}$$

$g(\boldsymbol{t})$ is called k times continuously differentiable if $\partial_{\boldsymbol{u}_1} \cdots \partial_{\boldsymbol{u}_\ell} g(\boldsymbol{t})$ exists, is continuous and is bounded for all $\boldsymbol{t} \in [0,1]^d$, $\ell \in \{0, 1, \ldots, k\}$ and $\boldsymbol{u}_i \in \mathbb{R}^d$ s.t. $|\boldsymbol{u}_i| = 1$, $i \in \{1, \ldots, \ell\}$.

For the letter $\boldsymbol{\ell}$, let $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_d)$ where $\ell_i \in \{0, 1, 2 \ldots\}$, and let $|\boldsymbol{\ell}| = \sum_j \ell_j$. For $\boldsymbol{h} \in \mathbb{R}^d$, let $\boldsymbol{h}^{\boldsymbol{\ell}} = h_1^{\ell_1} \cdots h_d^{\ell_d}$. For any function G from $\mathbb{R}^d$ to $\mathbb{R}$ let

$$G^{(\boldsymbol{\ell})} = \frac{\partial^{\ell_d}}{\partial t_d^{\ell_d}} \cdots \frac{\partial^{\ell_1}}{\partial t_1^{\ell_1}} G,$$

where $\frac{\partial^0}{\partial t_j^0}$ is interpreted as 1. For $F, G : \mathbb{R}^d \to \mathbb{R}$, if $|F(\boldsymbol{h})/G(\boldsymbol{h})| \leq c$ for some finite $c > 0$ and $|\boldsymbol{h}|$ sufficiently large or small then write $F(\boldsymbol{h}) = O(G(\boldsymbol{h}))$. Then if $|F(\boldsymbol{h})/G(\boldsymbol{h})| \leq c$ for all $c > 0$ and sufficiently large or small $|\boldsymbol{h}|$, write $F(\boldsymbol{h}) = o(\boldsymbol{h})$.

By Taylor's theorem, if g is k times continuously differentiable, then as $|\boldsymbol{h}| \to 0$,

$$g(\boldsymbol{t} + \boldsymbol{h}) = g(\boldsymbol{t}) + \sum_{\boldsymbol{\ell}: |\boldsymbol{\ell}| \leq k} \frac{1}{|\boldsymbol{\ell}|!} g^{(\boldsymbol{\ell})}(\boldsymbol{t}) \boldsymbol{h}^{\boldsymbol{\ell}} + R_{\boldsymbol{t}}(\boldsymbol{h})$$

where $R_{\boldsymbol{t}}(\boldsymbol{h}) = o(|\boldsymbol{h}|^k)$ uniformly over $\boldsymbol{t} \in [0,1]^d$. Note that uniform over $\boldsymbol{t} \in [0,1]^d$ means that $\sup_{\boldsymbol{t} \in [0,1]^d} |R_{\boldsymbol{t}}(\boldsymbol{h})| = o(|\boldsymbol{h}|^k)$. If $R_{\boldsymbol{t}}(\boldsymbol{h}) = O(|\boldsymbol{h}|^k)$ uniformly over $\boldsymbol{t} \in [0,1]^d$, then there is a single finite constant $c > 0$ s.t. $|R_{\boldsymbol{t}}(\boldsymbol{h})| \leq c|\boldsymbol{h}|^k$ for all $\boldsymbol{t} \in [0,1]^d$ and $|\boldsymbol{h}|$ sufficiently small.

For a function $g : \mathbb{R}^d \to \mathbb{R}$, define the gradient $\nabla g(\boldsymbol{t}) := \left(\frac{\partial}{\partial t_1} g(\boldsymbol{t}), \ldots, \frac{\partial}{\partial t_d} g(\boldsymbol{t}) \right)'$ and the Hessian matrix $H_g^{\boldsymbol{t}}$ whose value at row i column j is $\frac{\partial}{\partial t_i} \frac{\partial}{\partial t_j} g(\boldsymbol{t})$ for $i, j \in \{1, \ldots, d\}$. For a function $g : \mathbb{R}^d \to \mathbb{R}^d$ with $g = (g_1(\boldsymbol{t}), \ldots, g_d(\boldsymbol{t}))$, let $J_g^{\boldsymbol{t}}$ denote the Jacobian of g at $\boldsymbol{t}$ be the $d \times d$ matrix defined as $\frac{\partial}{\partial t_j} g_i(\boldsymbol{t})$ for row i and column j.

1.3 Stationarity

The simplest setting assumes that the covariance between two locations $\boldsymbol{t}$ and $\boldsymbol{s}$, denoted as $C(\boldsymbol{t}, \boldsymbol{s})$, only depends on the Euclidean distance $|\boldsymbol{t} - \boldsymbol{s}|$. In this setting Y is called isotropic. In this case, consistent estimation of the covariance parameters used in Kriging may be achieved even if the data set is only moderately sized.

Y is called stationary if $C(\boldsymbol{t}, \boldsymbol{s})$ depends on the distance and direction of $\boldsymbol{t} - \boldsymbol{s}$. In this setting we will need to estimate $C(\boldsymbol{t}, \boldsymbol{t} + \boldsymbol{h})$ for at least a few directions $\boldsymbol{h} = (h_1, \ldots, h_d)$ which will require either a structured data set or a dense sampling so that a sufficient number of pairs in our data are $\boldsymbol{h}$ apart. Y is called second order stationary if in addition μ is constant. Second order stationarity provides a convenient framework for analysis, but is often an unreasonable assumption to make for spatial processes.

The most complicated setting is when $C(\boldsymbol{t}, \boldsymbol{s})$ depends on the location $\boldsymbol{t}$ and changes across the domain. In this setting Y is called nonstationary and consistent estimation of $f_{\boldsymbol{t}}(\boldsymbol{h})|\boldsymbol{h}|^{\alpha}$ will require a dense sampling.

A reasonable assumption to make is that Y is locally stationary. This is when the covariance is approximately stationary in a local region or when $f_{\boldsymbol{t}}(\boldsymbol{h})$ is smoothly changing in $\boldsymbol{t}$.

If Y is locally stationary, one option would be to separate the domain into subdomains where the covariance structure is considered approximately stationary in each subdomain. Then estimation of the principal irregular term will require a sufficient number of points in each subdomain. Another option is to use a procedure like a moving window Kriging to estimate the principal irregular term at a location $\boldsymbol{t}_0$ with points that are within a window centered at $\boldsymbol{t}_0$ (assuming $\boldsymbol{t}_0$ is away from the boundary of the domain). With a procedure like moving window Kriging, estimation is aided considerably if the locations of the data are evenly spaced, or at least evenly spaced in each dimension. To that end, we assume that the domain has been sampled on a regular grid with spacing $1/n$ in each dimension and derive infill asymptotic convergence results as $n \to \infty$.

CHAPTER 2

Inference on Y

In this chapter we analyze the scenario where the observed process Y is assumed to be Gaussian and observed on a regular grid. Since Y is Gaussian, the mean $\mu(\boldsymbol{t})$ and the covariance $C(\boldsymbol{t}, \boldsymbol{s})$ characterize the behavior of Y. We assume μ is smooth and make a locally stationary assumption for C which are sufficient to recover the principal irregular term.

We present the Matérn covariance function, the deformation model, multifractional Brownian motion and intrinsic random functions as examples which satisfy our covariance assumptions. Then we examine linear prediction of Y through Kriging and justify an estimator which uses the principal irregular term. Lastly, we present the increment operator which we will use to estimate the principal irregular term.

2.1 Assumptions

Assumption 2.1. *The spatial process of interest is denoted as $Y(\boldsymbol{t})$, $\boldsymbol{t} \in \mathbb{R}^d$ for integer $d > 0$, and is assumed to be a Gaussian process with mean function μ and covariance function C. We observe the process on a regular grid*

$$T_n = \{(i_1, i_2, \ldots, i_d)/n, \text{ with } i_s = (j - 1/2)/n \text{ for } j = 1, \ldots, n\}$$

of size $1/n$ on $[0,1]^d$.

As mentioned in the introduction, we will also make assumptions on the local behavior of μ and C.

Assumption 2.2. *$\mu(\boldsymbol{t})$ is k times continuously differentiable for some sufficiently large integer k.*

Let $\partial_{\boldsymbol{u}}^{(i,j)} C(\boldsymbol{t}, \boldsymbol{s})$ denote the directional derivative in the direction $\boldsymbol{u}$ acting on $\boldsymbol{t}$ i times and acting on $\boldsymbol{s}$ j times.

Assumption 2.3. *For the covariance function C,*

(i) As $|\boldsymbol{t}-\boldsymbol{s}| \to 0$,

$$C(\boldsymbol{t},\boldsymbol{s}) = \sum_{|\boldsymbol{\ell}|=0}^{r} b_{\boldsymbol{\ell}}(\boldsymbol{t})\boldsymbol{s}^{\boldsymbol{\ell}} + \sum_{|\boldsymbol{\ell}|=0}^{r} b_{\boldsymbol{\ell}}(\boldsymbol{s})\boldsymbol{t}^{\boldsymbol{\ell}} + f_{\boldsymbol{t}}(\boldsymbol{t}-\boldsymbol{s})|\boldsymbol{t}-\boldsymbol{s}|^{\alpha(\boldsymbol{t})}(1+O(|\boldsymbol{t}-\boldsymbol{s}|^{\gamma(\boldsymbol{t})}),$$

where $f_{\boldsymbol{t}}(\boldsymbol{h})$ is sufficiently smooth in $\boldsymbol{t}$ and in $\boldsymbol{h}$, $\alpha(\boldsymbol{t})$ is continuously differentiable, $\alpha(\boldsymbol{t})/2$ is non-integer, $\alpha(\boldsymbol{t}) - r < 1$, $b_{\boldsymbol{\ell}}(\boldsymbol{t})$ are measurable functions, $\gamma(\boldsymbol{t}) > 0$ and $O(|\boldsymbol{t}-\boldsymbol{s}|^{\gamma(\boldsymbol{t})})$ is uniform over $\boldsymbol{t} \in [0,1]^d$.

(ii) $C(\boldsymbol{t},\boldsymbol{s})$ is x times continuously differentiable in $\boldsymbol{t}$ and $\boldsymbol{s}$ for $|\boldsymbol{t}-\boldsymbol{s}|$ away from 0 and $|\partial_{\boldsymbol{u}_1}^{(1,1)} \cdots \partial_{\boldsymbol{u}_x}^{(1,1)} C(\boldsymbol{t},\boldsymbol{s})| \leq \mathbf{c}_2|\boldsymbol{t}-\boldsymbol{s}|^{\alpha(\boldsymbol{t})-2x}$, for $\boldsymbol{u}_i \in \mathbb{R}^d$, $|\boldsymbol{u}_i| = 1$, $\mathbf{c}_2 > 0$ and some integer $x \geq 1$.

The parameter $\alpha(\boldsymbol{t})$ is a measure of smoothness of Y, where sample paths of Y are $\lfloor \alpha(\boldsymbol{t})/2 \rfloor$ times mean square differentiable. The value $\alpha(\boldsymbol{t})/2$ is also referred to as the Hurst parameter and is a measure of long-term dependence in time series data or surface roughness in higher dimensional dimensional random processes. Note that $f_{\boldsymbol{t}}(\boldsymbol{h})$ only depends on $\boldsymbol{h}$ through its direction.

Assumption 2.3(ii) is required for consistent estimation of the principal irregular term, ensuring that the covariance between increments converges to 0 quickly as the distance between increments increases. Assuming the observations fall on a regular grid is also necessary to ensure that a sufficient number of pairs of observations are separated by a lag $\boldsymbol{u} \in \mathbb{Z}^d$. Combined with the other assumptions, we will be able to consistently estimate $f_{\boldsymbol{t}}(\boldsymbol{u})|\boldsymbol{u}|^{\alpha(\boldsymbol{t})}$ at any location $\boldsymbol{t} \in [0,1]^d$.

2.2 Example Processes

Here are some examples of processes which satisfy our covariance assumption.

Example 2.4 (Anisotropic Matérn covariance function)**.** *The Matérn covariance function is defined as*

$$C(\boldsymbol{t},\boldsymbol{s}) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\frac{|\boldsymbol{t}-\boldsymbol{s}|}{\rho}\right)^{\nu} K_{\nu}\left(\sqrt{2\nu}\frac{|\boldsymbol{t}-\boldsymbol{s}|}{\rho}\right),$$

where $\nu > 0$ is the shape parameter, $\rho > 0$ is the scale parameter, σ^2 is the variance, Γ is the Gamma function and K_ν is the modified Bessel function of the second kind. If Z is a Gaussian random field with covariance function C and M is a $d \times d$ transformation matrix, then the field $Y(\boldsymbol{t}) = Z(M\boldsymbol{t})$ is a transformed Gaussian Matérn random field *with covariance $C(M\boldsymbol{t}, M\boldsymbol{s})$ which is anisotropic when M is not an orthogonal matrix up to a proportionality constant.*

This covariance function satisfies our covariance assumption with $\alpha(\boldsymbol{t}) = 2\nu$, and $f_{\boldsymbol{t}}(\boldsymbol{h}) = c\left|M\frac{\boldsymbol{h}}{|\boldsymbol{h}|}\right|^{2\nu}$ for some $c > 0$ and $x > \nu + 1$.

The flexibility of this covariance function makes it a popular choice in spatial statistics. Anderes (2011) showed that when M satisfies some regularity conditions, it can be recovered by averaging squared increments. He also shows that the parameters of the Matérn can be consistently estimated when $d > 4$, the parameters can't be consistently estimated separately when $d < 4$, and it is unknown whether they can be estimated separately when $d = 4$.

Example 2.5 (Multifractional Brownian Motion (mBm))**.** *$B_H(\boldsymbol{t})$, $|\boldsymbol{t}| > 0$ is a centered Gaussian process with covariance function*

$$C(\boldsymbol{t}, \boldsymbol{s}) = D(H(\boldsymbol{t}), H(\boldsymbol{s}))\left(|\boldsymbol{t}|^{H(\boldsymbol{t})+H(\boldsymbol{s})} + |\boldsymbol{s}|^{H(\boldsymbol{t})+H(\boldsymbol{s})} - |\boldsymbol{t} - \boldsymbol{s}|^{H(\boldsymbol{t})+H(\boldsymbol{s})}\right)$$

where $H(\boldsymbol{t}) : \mathbb{R}^d_+ \to (0, 1)$ is the smoothness of B at location $\boldsymbol{t}$ and D is a known deterministic function.

mBm satisfies our covariance assumpton with $\alpha(\boldsymbol{t}) = 2H(\boldsymbol{t})$, $f_{\boldsymbol{t}}(\boldsymbol{h}) = D(H(\boldsymbol{t}), H(\boldsymbol{t}))$ for all $\boldsymbol{h}$ and $H(\boldsymbol{t})$ is three times continuously differentiable.

When $H(\boldsymbol{t})$ is constant B_H is called fractional Brownian motion. Notice that this process is nonstationary, but has stationary increments. This process is also self-similar, i.e. $B_H(a\boldsymbol{t}) = |a|^{H(\boldsymbol{t})}B_H(\boldsymbol{t})$. B_H exhibits long-range dependence when $H > 1/2$. Lastly, sample paths are almost nowhere differentiable.

Example 2.6 (Deformation Model)**.** *Let Z be an isotropic Gaussian random field on $\Omega \subset \mathbb{R}^2$ with covariance $C(\boldsymbol{t}, \boldsymbol{t}+\boldsymbol{h}) = \sigma^2 - |\boldsymbol{h}|^\alpha + O(|\boldsymbol{h}|^2)$ where $\alpha \in (0, 2)$. Then if F is a smooth one-to-one deformation function, $Y(\boldsymbol{t}) = Z(F(\boldsymbol{t}))$ is a deformed random field with covariance*

$$C(F(\boldsymbol{t}), F(\boldsymbol{t}+\boldsymbol{h})) = \sigma^2 - |F(\boldsymbol{t}) - F(\boldsymbol{t}+\boldsymbol{h})|^\alpha + O(|F(\boldsymbol{t}) - F(\boldsymbol{t}+\boldsymbol{h})|^2).$$

For this model $f_{\boldsymbol{t}}(\boldsymbol{h}) = |J_F^{\boldsymbol{t}}\boldsymbol{h}|^\alpha$, where $J_F^{\boldsymbol{t}}$ is the Jacobian of F at $\boldsymbol{t}$. In the two dimensional case, Anderes and Chatterjee (2009) show that when F satisfies some regularity conditions, it can be recovered by estimates of the semivariogram $v(\boldsymbol{t}, \boldsymbol{t} + \boldsymbol{h}) \approx |J_F^{\boldsymbol{t}}\boldsymbol{h}|^\alpha$ in the horizontal, vertical and diagonal directions.

Deformation models are a method of modeling nonstationarity which assumes that an observed process is actually a smooth transformation of an underlying isotropic process. This method was introduced by Sampson and Guttorp(1992) and they considered estimating the deformation with multiple obervations of a deformed field at sparse locations. Clerc and Mallat(2003) and Anderes and Stein(2008) considered estimating the deformation from a single observation of a deformed random field, but with a dense sampling. Anderes and Chatterjee (2009) prove consistent estimation for a class of deformation functions known as quasiconformal maps through the use of kernel averaged squared increments while observing the deformed process on a regular dense grid.

For the next example, recall $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_d) \in \mathbb{N}^d$, $|\boldsymbol{\ell}| = \sum_{j=1}^d \ell_j$ and $\boldsymbol{h}^{\boldsymbol{\ell}} = h_1^{\ell_1} \cdots h_d^{\ell_d}$.

Example 2.7 (Intrinsic Random Functions)**.** *Y is called an intrinsic random function of order k (IRF-k) if*

$$C(\boldsymbol{t}, \boldsymbol{s}) = \sum_{|\boldsymbol{\ell}|=0}^{k} b_{\boldsymbol{\ell}}(\boldsymbol{t}) \boldsymbol{s}^{\boldsymbol{\ell}} + \sum_{|\boldsymbol{\ell}|=0}^{k} b_{\boldsymbol{\ell}}(\boldsymbol{s}) \boldsymbol{t}^{\boldsymbol{\ell}} + K(\boldsymbol{t} - \boldsymbol{s}),$$

where K is called the generalized covariance, the $b_{\boldsymbol{\ell}}$ are arbitrary measurable functions and the summations are over all $\boldsymbol{\ell}$ with $|\boldsymbol{\ell}| = 0, \ldots, k$.

This process was introduced in Matheron (1964) and covers a wide class of covariance functions. By definition, IRF-k's are stationary after a $k+1$ step differencing. This includes intrinsically stationary processes (IRF-0's) which are stationary after one-step differencing with K as their semivariogram. When $K(\boldsymbol{t} - \boldsymbol{s}) = f_{\boldsymbol{t}}(\boldsymbol{t} - \boldsymbol{s})|\boldsymbol{t} - \boldsymbol{s}|^{\alpha(\boldsymbol{t})} \left(1 + |\boldsymbol{t} - \boldsymbol{s}|^{\gamma(\boldsymbol{t})}\right)$, this process satisfies our covariance assumption.

2.3 Linear Prediction

A natural problem to conisder is prediction of $Y(\boldsymbol{t}_0)$ for some unobserved location $\boldsymbol{t}_0$, denoted as $\hat{Y}(\boldsymbol{t}_0)$.

The simplest prediction method is the nearest neighbor method, where $\hat{Y}(\boldsymbol{t}_0) = Y(\boldsymbol{t}_i)$ with $\boldsymbol{t}_i$ is the closest observed point to $\boldsymbol{t}_0$. Another method would be to take the simple average or a weighted average of points that are closest to $\boldsymbol{t}_0$. In general, a linear estimator of $Y(\boldsymbol{t}_0)$ can be written as

$$\hat{Y}(\boldsymbol{t}_0) = \sum_{i=1}^{m} \lambda_i Y(\boldsymbol{t}_i)$$

where $\boldsymbol{t}_i$ are points close to $\boldsymbol{t}_0$ and $\lambda_i \in \mathbb{R}$. When the mean is known, the best linear unbiased estimator will be the one which minimizes the error in estimating $Y(\boldsymbol{t}_0)$. If we consider squared-error loss as in chapter 3 of Cressie (1993), then the estimator minimizes $E(Y(\boldsymbol{t}_0) - \hat{Y}(\boldsymbol{t}_0))^2$. This estimator is known as the Kriging estimator.

There are different types of Kriging which depend on the mean and covariance structure. The simplest type of Kriging occurs when the mean is known and constant and is called simple Kriging. Ordinary Kriging is used when the mean is unknown but assumed to be constant in a local neighborhood and the variogram is estimable from the data. Let $\lambda = (\lambda_1, \dots, \lambda_m)'$, then the coefficients for ordinary Kriging are

$$\boldsymbol{\lambda}' = \left(\boldsymbol{v} + \mathbf{1}\frac{(1 - \mathbf{1}'V^{-1}\boldsymbol{v})}{\mathbf{1}'V^{-1}\mathbf{1}}\right)' V^{-1},$$

where $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d)'$, $\boldsymbol{v} = (v(\boldsymbol{t}_0, \boldsymbol{t}_1), \dots, v(\boldsymbol{t}_0, \boldsymbol{t}_m))'$, $\mathbf{1}$ is a vector of ones, and V is the $m \times m$ matrix with $(i, j)^{th}$ element equal to $v(\boldsymbol{t}_i, \boldsymbol{t}_j)$. The most general forms of Kriging occur when the mean is unknown and non-constant. Universal Kriging assumes the mean can be modeled as a linear combination of functions $f_j(\boldsymbol{t})$, $j = 0, \dots, p$. This type leads to coefficients

$$\boldsymbol{\lambda}' = \left(\boldsymbol{v} + X(X'V^{-1}X)^{-1}(\boldsymbol{x} - XV^{-1}\boldsymbol{v})\right)' V^{-1},$$

where X is an $n \times (p+1)$ matrix with $(i, j)^{th}$ element is equal to $f_{j-1}(\boldsymbol{t}_i)$ and $\boldsymbol{x} = (f_0(\boldsymbol{t}_0), \dots, f_p(\boldsymbol{t}_0))'$. Let the universal Kriging predictor using the true variogram and assuming the mean is polynomial of degree k be denoted as $\hat{Y}_{TK-k}$.

Here we develop the Kriging estimator that is of most interest, which is closely related to IRF Kriging as outlined in Chilés and Delfiner (1999). For convenience, let $\lambda_0 = -1$. The best linear unbiased predictor $\hat{Y}$ can be obtained by minimizing

$$\begin{aligned} E(\hat{Y}(\boldsymbol{t}_0) - Y(\boldsymbol{t}_0))^2 &= \operatorname{Var}(\hat{Y}(\boldsymbol{t}_0) - Y(\boldsymbol{t}_0)) + E(\hat{Y}(\boldsymbol{t}_0) - Y(\boldsymbol{t}_0))^2 \\ &=: V_1 + V_2 + V_3 + V_4, \end{aligned}$$

where

$$\begin{aligned} V_1 &= 2\sum_{|\boldsymbol{\ell}|=0}^{r}\sum_{j=0}^{m}\lambda_j b_\ell(\boldsymbol{t}_j)\sum_{i=0}^{m}\lambda_i \boldsymbol{t}_i^{\boldsymbol{\ell}} \\ V_2 &= \sum_{i=0}^{m}\sum_{j=0}^{m}\lambda_i\lambda_j f(\boldsymbol{t}_i - \boldsymbol{t}_j)|\boldsymbol{t}_i - \boldsymbol{t}_j|^{\alpha(\boldsymbol{t}_i)} \\ V_3 &= \sum_{i=0}^{m}\sum_{j=0}^{m} O(|\boldsymbol{t} - \boldsymbol{t}_j|^{\alpha(\boldsymbol{t}_i)+\gamma(\boldsymbol{t}_i)}) \end{aligned}$$

$$V_4 = (\sum_{i=0}^{m} \lambda_i \mu(\boldsymbol{t}_i))^2.$$

From our sampling scheme, we can consider locations $\boldsymbol{t}_i$ s.t. $|\boldsymbol{t}_i - \boldsymbol{t}_0| = O(n^{-1})$. Since μ is $k+1$ times continuously differentiable we can write

$$\begin{aligned}\mu(\boldsymbol{t}_i) &= \mu(\boldsymbol{t}_0) + \sum_{j=1}^{d} \frac{\partial}{\partial t_j} \mu(\boldsymbol{t}_0)(t_{0j} - t_{ij}) + \ldots \\ &\quad + \sum_{j_1,\ldots,j_k=1}^{d} \frac{\partial}{\partial t_{j_1}} \cdots \frac{\partial}{\partial t_{j_k}} \mu(\boldsymbol{t}_0)(t_{0j_1} - t_{ij_1}) \cdots (t_{0j_k} - t_{ij_k}) + O(n^{-k-1}).\end{aligned}$$

If we add the constraint

$$\sum_{i=0}^{m} \lambda_i \boldsymbol{t}_i^{\boldsymbol{\ell}} = 0,\ \boldsymbol{\ell} \in \mathbb{N}^d \text{ s.t. } |\boldsymbol{\ell}| = 0, \ldots, k \tag{2.1}$$

to the optimization problem, then the low order terms in this expansion are eliminated. This constraint also ensures that $V_1 = 0$ if $k \geq r$. Therefore

$$E\{\hat{Y}(\boldsymbol{t}_0) - Y(\boldsymbol{t}_0)\}^2 \approx \sum_{i=0}^{m} \sum_{j=0}^{m} \lambda_i \lambda_j f(\boldsymbol{t}_i - \boldsymbol{t}_j) |\boldsymbol{t}_i - \boldsymbol{t}_j|^{\alpha(\boldsymbol{t}_i)}. \tag{2.2}$$

By Lagrange's multiplier, the problem of optimizing (2.2) under (2.1) becomes solving for the system

$$\begin{aligned}&\sum_{j=1}^{m} \lambda_j f(\boldsymbol{t}_i - \boldsymbol{t}_j)|\boldsymbol{t}_i - \boldsymbol{t}_j|^{\alpha} + \sum_{|\boldsymbol{\ell}|=0}^{r} \mu_{\boldsymbol{\ell}} \boldsymbol{t}_i^{\boldsymbol{\ell}} = f(\boldsymbol{t}_i - \boldsymbol{t}_0)|\boldsymbol{t}_i - \boldsymbol{t}_0|^{\alpha}, i = 1, \ldots, N, \\ &\sum_{j=0}^{m} \lambda_j \boldsymbol{t}_j^{\boldsymbol{\ell}} = \boldsymbol{t}_0^{\boldsymbol{\ell}},\ \ell_1, \ell_2, \ldots, \ell_d \geq 0,\ \ell_1 + \ell_2 + \ldots + \ell_d \leq k.\end{aligned}$$

We can express this in matrix form:

$$\begin{bmatrix} \boldsymbol{K} & \boldsymbol{X} \\ \boldsymbol{X}' & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\mu} \end{bmatrix} = \begin{bmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \end{bmatrix}$$

where

$$\begin{aligned}\boldsymbol{\lambda} &= (\lambda_1, \ldots, \lambda_m)', \\ \boldsymbol{\mu} &= \text{ a vector of } \mu_{\boldsymbol{\ell}} \text{ arranged in any order,}\end{aligned}$$

$$
\begin{aligned}
\boldsymbol{K} &= \{f(\boldsymbol{t}_i - \boldsymbol{t}_j)|\boldsymbol{t}_i - \boldsymbol{t}_j|^{\alpha}\}_{i,j=1}^{m}, \\
\boldsymbol{X} &= \text{ a matrix with } m \text{ rows where the } i\text{-th row is the vector containing} \\
& \quad \boldsymbol{t}_i^{\boldsymbol{\ell}}\text{; the order is arranged in the same way as } \boldsymbol{\mu}, \\
\boldsymbol{v}_1 &= (f(\boldsymbol{t}_1 - \boldsymbol{t}_0)|\boldsymbol{t}_1 - \boldsymbol{t}_0|^{\alpha}, \ldots, f(\boldsymbol{t}_N - \boldsymbol{t}_0)|\boldsymbol{t}_N - \boldsymbol{t}_0|^{\alpha})', \\
\boldsymbol{v}_2 &= \text{ the vetor containing } \boldsymbol{t}_0^{\boldsymbol{\ell}} \text{ arranged in the same order as } \boldsymbol{\mu}.
\end{aligned}
$$

The resulting coefficients for this estimator are

$$
\boldsymbol{\lambda}' = \left(\boldsymbol{v}_1 + \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{K}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{v}_2 - \boldsymbol{X}\boldsymbol{K}^{-1}\boldsymbol{v}_1)\right)' \boldsymbol{K}^{-1}
$$

and we call a predictor of this form the $UK-k$ estimator

$$
\hat{Y}_{UK-k} = \boldsymbol{\lambda}'\boldsymbol{Y}.
$$

Note that this will be the estimator using the estimated f.

By construction, the $TK-0$ estimator minimizes the estimation error if the mean is constant. When the mean is non-constant, a larger value of k may improve estimation by eliminating the effect of the mean. But if the mean is constant, the $TK-k$ estimators for $k \geq 1$ will not have smaller estimation errors than the $TK-0$ estimator. Similarly, adding in the linear constraints of (2.1) for the $UK-k$ estimator won't result in a reduction of estimation error over the $TK-0$ estimator. However, the differencing constraints are necessary for the $UK-k$ estimators to minimize the variance when the principal irregular term is used.

2.4 Inference on the Principal Irregular Term

Recall from section 1.2 that $\Delta_{\boldsymbol{h}/n}^{x} Y(\boldsymbol{t})$ is the increment operator of order x in the direction $\boldsymbol{h}/n$. Since the sampling grid is regular with spacing $1/n$, the $\boldsymbol{h}$ we can consider are $\boldsymbol{h} = (h_1, \ldots, h_d)$ where h_i are integers. For now assume that $\alpha(\boldsymbol{t})$ is known and consider the process

$$
W_n(\boldsymbol{t}) = n^{\alpha(\boldsymbol{t})/2} \Delta_{\boldsymbol{h}/n}^{x} Y(\boldsymbol{t})
$$

for some integer x. Then let

$$
J(\boldsymbol{t}, \boldsymbol{u}) = \sum_{i=0}^{x} \sum_{j=0}^{x} (-1)^{i+j} \binom{x}{i} \binom{x}{j} f_{\boldsymbol{t}}(\boldsymbol{u} + (i-j)\boldsymbol{h})|\boldsymbol{u} + (i-j)\boldsymbol{h}|^{\alpha(\boldsymbol{t})}.
$$

Assumption 2.3 leads to the following lemma for W_n.

Lemma 2.8. *Let Assumption 2.3 hold with $x > r$.*

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

$$\mathrm{Cov}(W_n(\boldsymbol{t}), W_n(\boldsymbol{t}+\boldsymbol{u}/n)) - J_{\boldsymbol{t}}(\mathbf{u}) = O(n^{-1}\log(n)) + O(n^{-\gamma(\boldsymbol{t})}),$$

uniformly over $\boldsymbol{t} \in [0,1]^d$ where $O(n^{-1}log(n))$ is replaced by $O(n^{-1})$ if $\alpha(\boldsymbol{t})$ is constant.

(ii) For any fixed $\boldsymbol{h}$, there exists $\mathbf{c}_4 > 0$ s.t.

$$\Big|\mathrm{Cov}(W_n(\boldsymbol{t}), W_n(\boldsymbol{t}+\boldsymbol{u}/n))\Big| \leq \mathbf{c}_4|\boldsymbol{u}|^{\alpha(\boldsymbol{t})-2x}$$

for $|\boldsymbol{u}| > \frac{x|\boldsymbol{h}|+1}{n}$.

Note that the proofs of this result and the results of the next chapter are provided in appendix II. Combining Assumption 2.2 and lemma 2.8, it seems possible to estimate $J(\boldsymbol{t},\boldsymbol{h})$ by averaging $W^2(\boldsymbol{t}_i)$ for $\boldsymbol{t}_i$ within a small window of $\boldsymbol{t}_0$.

In the next chapter, we will introduce the local linear kernel regression estimator of $J(\boldsymbol{t},\boldsymbol{h})$. We will prove a central limit theorem and a uniform convergence rate.

CHAPTER 3

Limit Theorems

In this chapter we introduce the local linear kernel regression estimator. We consider the abstract Gaussian process $W_n(\boldsymbol{t})$ introduced in the last chapter whose variance is approximately $J(\boldsymbol{t}, \mathbf{0})$. The local linear kernel regression is a weighted average of $W_n^2(\boldsymbol{t}_i)$ estimating $J(\boldsymbol{t}_0, \mathbf{0})$, where $\boldsymbol{t}_i$ are the spatial locations within the d-dimensional square centered at $\boldsymbol{t}_0$.

We make some basic assumptions on the mean and covariance of the process W_n and the kernel function k. With these assumptions we prove the local linear smoother is consistent by showing the bias of the smoother convergences to 0 (Theorem 3.3) and the variance converges to 0 (Theorem 3.4). Then Theorem 3.5 gives a central limit theorem for the smoother. Lastly, assuming that k has a product form, Theorem 3.8 proves an almost sure uniform convergence result.

We focus on the abstract Gaussian process W_n with mean μ_n and covariance C_n.

Assumption 3.1. *For μ_n and C_n:*

(i) There exists a positive, continuously differentiable function $\delta(\boldsymbol{t})$ s.t. $\mu_n(\boldsymbol{t}) = O(n^{-\delta(\boldsymbol{t})})$ uniformly in $\boldsymbol{t}$.

(ii) There exists a positive, continuously differentiable function $\rho(\boldsymbol{t})$ and a twice continuously differentiable function $g(\boldsymbol{t})$ such that $C_n(\boldsymbol{t}, \boldsymbol{t}) = g(\boldsymbol{t}) + O(n^{-\rho(\boldsymbol{t})})$ uniformly in $\boldsymbol{t}$ with $\rho(\boldsymbol{t}) > 0$.

(iii) $C_n(\boldsymbol{t}, \boldsymbol{t} + \boldsymbol{h}/n) < \mathbf{c}_5 \ \forall \ \boldsymbol{t}, n, \boldsymbol{h}$ and $\mathbf{c}_5 > 0$ and there exists a function $g_2(\boldsymbol{t}, \boldsymbol{h})$ such that $\lim_{n\to\infty} C_n(\boldsymbol{t}, \boldsymbol{t} + \boldsymbol{h}/n) = g_2(\boldsymbol{t}, \boldsymbol{h})$. The convergence is uniform for all $\boldsymbol{t} \in [0,1]^d$ and all $\boldsymbol{h}$ with $|\boldsymbol{h}| \leq \delta$ for any given $\delta > 0$.

(iv) There exists some function $\psi(\boldsymbol{t})$ with values in $(0, \infty)$ and positive constant $\mathbf{c}_6$ such that $|C_n(\boldsymbol{t}, \boldsymbol{t} + \boldsymbol{h}/n)| \leq \mathbf{c}_6 |\boldsymbol{h}|^{-\psi(\boldsymbol{t})}$ for all $n, \boldsymbol{t}, \boldsymbol{h}$ such that $|\boldsymbol{h}| > \tau$ for some $\tau > 0$.

Note that uniformly in $\boldsymbol{t}$ means that the constants in the big-o and little-o terms do not depend on $\boldsymbol{t}$.

Notice that by lemma 2.8, W_n satisfies Assumption 3.1 with $\rho(\boldsymbol{t}) < \min(1, \gamma(\boldsymbol{t}), 2(m+1) - \alpha(\boldsymbol{t}))$, where $<$ is changed to $=$ if $\alpha(\boldsymbol{t})$ is constant.

Let

$$T_n(\boldsymbol{t}_0, b) = \{\boldsymbol{t}_i \in T_n \cap [\boldsymbol{t}_0 - b, \boldsymbol{t}_0 + b]\}$$

where

$$[\boldsymbol{t}_0 - b, \boldsymbol{t}_0 + b] := \prod_{j=1}^{d} [t_{0j} - b, t_{0j} + b] .$$

Throughout the rest of the thesis, let $k(\cdot)$ denote a kernel function that satisfies the following assumption.

Assumption 3.2. *Assume $k(\cdot)$ is a non-negative function with support $[-1,1]^d$ and has bounded, continuous first order partial derivatives with $\|k\|_\infty = \sup_{\boldsymbol{t}\in[-1,1]^d} k(\boldsymbol{t})$.*

For any $\boldsymbol{t}_0 \in [0,1]^d$ and bandwidth b, the local linear regression is defined as

$$\hat{\beta}(\boldsymbol{t}_0; n, b) = \text{argmin}_\beta \sum_{\boldsymbol{t}_i \in T_n(\boldsymbol{t}_0, b)} k\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right) \{W_n^2(\boldsymbol{t}_i) - \ell(\boldsymbol{t}_i - \boldsymbol{t}_0; \beta)\}^2$$

where

$$\ell(\boldsymbol{t}; \beta) = \beta_0 + \beta_1 t_1 + \beta_2 t_2 + \cdots + \beta_d t_d \text{ for } \boldsymbol{t} = (t_1, t_2, \ldots, t_d).$$

It follows that

$$\hat{\beta}(\boldsymbol{t}_0; n, b) = (X'KX)^{-1} X'KW^2,$$

where X is a matrix with rows $(1, t_{i1} - t_{01}, t_{i2} - t_{02}, \ldots, t_{id} - t_{0d})$, K is a diagonal matrix with entries $k\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right)$, and W^2 is a column vector with entries $(W_n(\boldsymbol{t}_i))^2$. For $\boldsymbol{t} \in [0,1]^d$, let

$$a_{i1} = \frac{t_i}{b} \wedge 1, \quad a_{i2} = \tfrac{1-t_i}{b} \wedge 1.$$

Define

$$I(\boldsymbol{t}, b) \quad = \quad [-a_{11}, a_{12}] \times [-a_{21}, a_{22}] \times \cdots \times [-a_{d1}, a_{d2}]$$

$$= \frac{\left(([t_1 - b, t_1 + b] \times \cdots \times [t_d - b, t_d + b]) \cap [0,1]^d\right) - (t_1, \ldots, t_d)}{b}.$$

If the distance between $\boldsymbol{t}$ and the boundary of $[0,1]^d$ is at least b then $a_{ij} = 1$ for all i, j, but some a_{ij} can be less than 1 if $\boldsymbol{t}$ is within distance b from some boundary point. For example, for $\boldsymbol{t} = (0,0)$ and $b \leq 1$, we have $a_{11} = a_{21} = 0$ and $a_{12} = a_{22} = 1$. Let

$$\begin{aligned}
S_{1^{m_1},\ldots,d^{m_d}} &= \sum_{\boldsymbol{t}_i} k\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right)(t_{i1} - t_{01})^{m_1} \cdots (t_{id} - t_{0d})^{m_d}, \\
\kappa_{1^{m_1},\ldots,d^{m_d}} &= \kappa_{1^{m_1},\ldots,d^{m_d}}(\boldsymbol{t}_0, b) = \int\limits_{I(\boldsymbol{t}_0,b)} k(\boldsymbol{z}) z_1^{m_1} \cdots z_d^{m_d} d\boldsymbol{z}.
\end{aligned}$$

For simplicity, if $m_j = 0$ for dimension j, then leave j^{m_j} out of $S_{1^{m_1},\ldots,d^{m_d}}$ and $\kappa_{1^{m_1},\ldots,d^{m_d}}$. For example,

$$S_{1^2,3^1} = \sum_{\boldsymbol{t}_i} k\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right)(t_{i1} - t_{01})^2 (t_{i3} - t_{03}).$$

From Taylor's theorem,

$$g(\boldsymbol{t}_i) = g(\boldsymbol{t}_0) + (\boldsymbol{t}_i - \boldsymbol{t}_0)' \nabla g(\boldsymbol{t}_0) + \frac{1}{2}(\boldsymbol{t}_i - \boldsymbol{t}_0)' H_g(\boldsymbol{t}_0)(\boldsymbol{t}_i - \boldsymbol{t}_0) + O(|\boldsymbol{t}_i - \boldsymbol{t}_0|^3),$$

for gradient $\nabla g(\boldsymbol{t})$ and Hessian matrix H_g. Therefore, $\hat{\beta}(\boldsymbol{t}_0; n, b)$ estimates

$$\left(g(\boldsymbol{t}_0), \frac{\partial}{\partial t_1} g(\boldsymbol{t}_0), \frac{\partial}{\partial t_2} g(\boldsymbol{t}_0), \ldots, \frac{\partial}{\partial t_d} g(\boldsymbol{t}_0)\right) := \beta(\boldsymbol{t}_0).$$

And further the bias terms of our estimator will involve kernel averages of $(\boldsymbol{t}_i - \boldsymbol{t}_0)$ terms and derivatives of g, so we define

$$\mathcal{K} = \begin{pmatrix}
\kappa & \kappa_{1^1} & \kappa_{2^1} & \cdots & \kappa_{d^1} \\
\kappa_{1^1} & \kappa_{1^2} & \kappa_{1^1,2^1} & \cdots & \kappa_{1^1,d^1} \\
\kappa_{2^1} & \kappa_{1^1,2^1} & \kappa_{2^2} & \cdots & \kappa_{2^1,d^1} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\kappa_{d^1} & \kappa_{1^1,d^1} & \kappa_{2^1,d^1} & \cdots & \kappa_{d^2}
\end{pmatrix}, \mathcal{N}_{i,j} = \begin{pmatrix}
\kappa_{i^1 j^1} \\
\kappa_{i^1 j^1 1^1} \\
\kappa_{i^1 j^1 2^1} \\
\vdots \\
\kappa_{i^1 j^1 d^1}
\end{pmatrix}.$$

Then we have the following theorem for the bias of $\hat{\beta}$.

Theorem 3.3 (Bias of $\hat{\beta}$)**.** *Let Assumptions 3.1(i) and (ii) hold for dimension d.*

Then if $n \to \infty$, $nb \to \infty$,

$$\mathbb{E}(\hat{\beta}(\boldsymbol{t}_0; n, b)) = \beta(\boldsymbol{t}_0) + \frac{\operatorname{diag}(b^2, b, b, \ldots, b)}{2} \mathcal{K}^{-1} \sum_{i,j=1}^{d} \frac{\partial}{\partial t_i} \frac{\partial}{\partial t_j} g(\boldsymbol{t}_0) \mathcal{N}_{i,j} + \begin{pmatrix} o(b^2) + O(n^{-(\rho(\boldsymbol{t}_0) \wedge 2\delta(\boldsymbol{t}_0))}) \\ o(b) + b^{-1} O(n^{-(\rho(\boldsymbol{t}_0) \wedge 2\delta(\boldsymbol{t}_0))}) \\ \vdots \\ o(b) + b^{-1} O(n^{-(\rho(\boldsymbol{t}_0) \wedge 2\delta(\boldsymbol{t}_0))}) \end{pmatrix} \quad (3.1)$$

uniformly for $\boldsymbol{t}_0 \in [0,1]^d$.

From this theorem we see that the bias of $\beta_0(\boldsymbol{t}_0)$ is $O(b^2) + O(n^{-(\rho(\boldsymbol{t}_0) \wedge 2\delta(\boldsymbol{t}_0))})$.

Let $\widetilde{W}_n(\boldsymbol{t}_i) = W_n(\boldsymbol{t}_i) - \mu_n(\boldsymbol{t}_i)$ then define

$$\tilde{\beta}(\boldsymbol{t}_0; n, b) := (X'KX)^{-1} X'K\widetilde{W}^2,$$

where $\widetilde{W}^2$ is a vector with entries $\widetilde{W}_n^2(\boldsymbol{t}_i)$. Define

$$\begin{aligned} \bar{k}(\boldsymbol{z}) &= k(\boldsymbol{z}) \Big(\mathcal{K}_{1,1}^{-1} + \sum_{j=1}^{d} \mathcal{K}_{1,1+j}^{-1} z_j \Big), \\ A(\boldsymbol{t}) &= 2 \int_{I(\boldsymbol{t}, b)} \bar{k}(\boldsymbol{z})^2 d\boldsymbol{z} \sum_{\boldsymbol{j} \in \mathbb{Z}^d} g_2^2(\boldsymbol{t}, \boldsymbol{j}). \end{aligned}$$

Note that, by Fatou's lemma and Assumption 3.1(iv),

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

for some finite constant C.

Recall the definition of the "double factorial"

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

In the formulation of the central limit theorem we need to bound quantities of the form $|\sum_i C_n(\boldsymbol{t}, \boldsymbol{t}_i)|$ (lemma B.5), which by Assumptions 3.1(ii)-(iv) has the rate

$$r_{nb}(\boldsymbol{t}) := \begin{cases} 1 & \text{if } \psi(\boldsymbol{t}) > d \\ \log(nb) & \text{if } \psi(\boldsymbol{t}) = d \\ (nb)^{d - \psi(\boldsymbol{t}_0)} & \text{if } \psi(\boldsymbol{t}) < d. \end{cases}$$

Then the following theorem establishes the behavior of the central moments of $\tilde{\beta}_0$.

Theorem 3.4 (x^{th} central moment of $\hat{\beta}_0(\boldsymbol{t}_0)$)**.** *Let Assumption 3.1 hold. Then for* $x = 2, 3, \ldots,$

$$(nb)^{dx/2} E\left(\frac{\tilde{\beta}_0(\boldsymbol{t}_0; n, b) - E\tilde{\beta}_0(\boldsymbol{t}_0; n, b)}{A(\boldsymbol{t}_0, b)^{x/2}}\right)^x = \begin{cases} O\left(\frac{r_n b(\boldsymbol{t}_0)}{(nb)^{d/2}}\right), & \text{for } x \text{ odd} \\ (x-1)!! + o(1), & \text{for } x \text{ even} \end{cases} \tag{3.2}$$

uniformly for $\boldsymbol{t}_0 \in [0,1]^d$ *as* $nb \to \infty$.

Note that the right-hand side of (3.2) converges to the x^{th} moments of the standard normal distribution. Then by Theorem B.6, this along with Theorem 3.3 is sufficient to establish Theorem 3.5.

Theorem 3.5. *[CLT for $\hat{\beta}_0$] Let Assumptions 3.1 and 3.2 hold with $\psi(\boldsymbol{t}_0) > d/2$. Then for $\boldsymbol{t}_0 \in [0,1]^2$, $Z \sim N(0,1)$,*

$$\hat{\beta}_0(\boldsymbol{t}_0; n, b) \stackrel{d}{=} g(\boldsymbol{t}_0) + \frac{ZA(\boldsymbol{t}_0, b)^{1/2}}{(nb)^{d/2}} + \frac{1}{2} b^2 [\mathcal{K}^{-1} \sum_{i,j=1}^{d} \frac{\partial}{\partial t_i} \frac{\partial}{\partial t_j} g(\boldsymbol{t}_0) \mathcal{N}_{i,j}]_1 + R(\boldsymbol{t}_0; n, b),$$

where

$$R(\boldsymbol{t}; n, b) = o(b^2) + O\left(n^{-(\rho(\boldsymbol{t}_0) \wedge 2\delta(\boldsymbol{t}_0))}\right) + O_p\left(n^{-\delta(\boldsymbol{t}_0)} \sqrt{r_{nb}(\boldsymbol{t})}\right) + o\left((nb)^{-d/2}\right).$$

Furthermore, when $n \to \infty$ and $nb \to \infty$, we can choose b s.t.

$$A(\boldsymbol{t}_0, b)^{-1/2} (nb)^{d/2} \left\{\hat{\beta}_0(\boldsymbol{t}_0; n, b) - g(\boldsymbol{t}_0)\right\} \rightarrow Z.$$

Note that this is the leads to the optimal rate for estimating a regression function nonparametrically assuming two derivatives and i.i.d. errors.

So far we have considered the asymptotic behavior of $\hat{\beta}_0(\boldsymbol{t}; n, b)$ for a fixed $\boldsymbol{t}$. Next, we consider the global asymptotic behavior of $\hat{\beta}_0(\boldsymbol{t}; n, b)$. In our proof of a.s. convergence, our estimators have the form

$$\frac{1}{n^d b^d} \sum_{\boldsymbol{t}_i \in T_n(\boldsymbol{t}_0, b)} k\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right) W_n(\boldsymbol{t}_i)^2 \tag{3.3}$$

and we will require the following conditions for k.

Assumption 3.6 (Almost sure convergence conditions for k)**.** *Assume for functions k, k_i, $i = 1, \ldots, d$,* *(i)* k *has support on* $[-1,1]^d$ *and* k_i *have support on* $[-1,1]$,

(ii) k_i *are defined s.t.* $k(\boldsymbol{z}) = \prod_{i=1}^d k_i(z_i)$,

(iii) k_i *are continuously differentiable and* $k_i(-1) = 0$ *for all* $i = 1, 2, \ldots, d$.

In addition, our proof of a.s. convergence does not restrict the value of x. To that end, define

$$s_{nb}(\boldsymbol{t}) := \begin{cases} 1 & \text{if } \psi(\boldsymbol{t}) > d/2 \\ \log(nb) & \text{if } \psi(\boldsymbol{t}) = d/2 \\ (nb)^{d-2\psi(\boldsymbol{t}_0)} & \text{if } \psi(\boldsymbol{t}) < d/2. \end{cases}$$

Then we have the following is a corollary to Theorem 3.4

Corollary 3.7. *Let Assumptions 3.1(iii) and (iv) hold and let* $[\boldsymbol{u}, \boldsymbol{v}]$ *denote the rectangle with corners* $\boldsymbol{u}$ *and* $\boldsymbol{v}$*. Then for sufficiently large* n*, there is a constant* $\mathbf{c}_7 > 0$ *such that*

$$\frac{(nb)^{-dx/2}}{s_{nb}(\boldsymbol{t})^{x/2}} E\left\{ \sum_{\boldsymbol{t}_i \in [\boldsymbol{u},\boldsymbol{v}] \cap T_n} (W_n(\boldsymbol{t}_i) - \mu_n(\boldsymbol{t}_i))^2 - E(W_n(\boldsymbol{t}_i) - \mu_n(\boldsymbol{t}_i))^2 \right\}^x \leq (2x-1)!!\mathbf{c}_7^x,$$

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

Let

$$V_{n,b} = \max\left\{1, \max_{\boldsymbol{t} \in [0,1]^d} s_{nb}(\boldsymbol{t})^{1/2}, \max_{\boldsymbol{t} \in [0,1]^d} r_{nb}(\boldsymbol{t})^{1/2} n^{-\delta(\boldsymbol{t})}, (nb)^{d/2}(\log(n))^{-1}(b^2 + n^{-\rho \wedge 2\delta})\right\},$$

where $\rho = \min_{\boldsymbol{t} \in [0,1]^d} \rho(\boldsymbol{t})$ and $\delta = \min_{\boldsymbol{t} \in [0,1]^d} \delta(\boldsymbol{t})$. Then we have the following a.s. uniform convergence result.

Theorem 3.8. *Let Assumptions 3.1(i)-(iv) and 3.6 hold and let* n *and* b *satisfy* $n \to \infty$ *and* $nb \to \infty$*. Then for some finite* $\mathbf{c}_8 > 0$,

$$\sup_{\boldsymbol{t}_0 \in [0,1]^2} \left| \hat{\beta}_0(\boldsymbol{t}_0; n, b) - g(\boldsymbol{t}_0) \right| \leq \mathbf{c}_8 (nb)^{-d/2} \log(n) V_{n,b}$$

eventually with probability 1.

Corollary 3.9. *Let Assumptions 3.1(i)-(iv) and 3.6 hold and let* n *and* b *satisfy* $n \to \infty$ *and* $nb \to \infty$*. Then for some finite* $\mathbf{c}_9 > 0$ *and* $i = 1, \ldots, d$,

$$\sup_{\boldsymbol{t}_0 \in [0,1]^2} \left| \hat{\beta}_i(\boldsymbol{t}_0; n, b) - \frac{\partial}{\partial t_i} g(\boldsymbol{t}_0) \right| \leq \mathbf{c}_9 (nb)^{-d/2} b^{-1} \log(n) V_{n,b}$$

eventually with probability 1.

CHAPTER 4

Related Issues

In this chapter we discuss issues related to the local linear smoother presented in the last chapter. First we give a theoretical calculation and plot to show that the local linear estimator estimates better at the boundary than the Nadaraya-Watson estimator. Then we show that the convergence rate of the local linear smoother matches the optimal nonparametric rate from Stone (1982) and the uniform rate is slightly less than optimal. Then we show the rate in estimating α is similar to the rate in estimating f with α known, and further that the rate in estimating f and α simultaneously worsens by $\log(n)$. Next, we give justification that the the lowest order difference which satisfies our assumption is optimal. Lastly, we give a brief treatment for what happens when data is missing at random.

4.1 Boundary Points

One reason we consider the local linear estimator is because it will improve the bias when estimating close to the boundary. If we examine the Nadaraya-Watson estimator for a point $\boldsymbol{t}_0$ away from the boundary, recalling $S_{00} = \sum_{\boldsymbol{t}_i} k\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right)$ and the results of chapter 3,

$$
\begin{aligned}
& S_{00}^{-1} \sum_{\boldsymbol{t}_i} k\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right) \mathbb{E}(\Delta^x_{\boldsymbol{h}/n} Y(\boldsymbol{t}_i))^2 \\
\approx \; & S_{00}^{-1} \sum_{\boldsymbol{t}_i} k\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right) g(\boldsymbol{t}_i) \\
= \; & S_{00}^{-1} \sum_{\boldsymbol{t}_i} k\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right) \Big\{ g(\boldsymbol{t}_0) + (t_{i_1} - t_{01}) \frac{\partial}{\partial t_1} g(\boldsymbol{t}_0) \\
& \qquad + (t_{i_2} - t_{02}) \frac{\partial}{\partial t_2} g(\boldsymbol{t}_0) + O\left(|\boldsymbol{t}_i - \boldsymbol{t}_0|^2\right) \Big\}
\end{aligned}
$$

$$
\begin{aligned}
&= g(\boldsymbol{t}_0) + S_{00}^{-1}S_{10}\frac{\partial}{\partial t_1}g(\boldsymbol{t}_0) + S_{00}^{-1}S_{01}\frac{\partial}{\partial t_2}g(\boldsymbol{t}_0) + O(b^2) \qquad (4.1)\\
&= g(\boldsymbol{t}_0) + O(b^2)
\end{aligned}
$$

Since S_{10} and S_{01} equal 0 when $\boldsymbol{t}_0$ is away from the boundary. So asymptotically, the N-W estimator has the same bias rate as the local linear estimator. However, if $\boldsymbol{t}_0$ is close to the boundary, then (4.1) becomes

$$
\begin{aligned}
& g(\boldsymbol{t}_0) + S_{00}^{-1}S_{10}\frac{\partial}{\partial t_1}g(\boldsymbol{t}_0) + S_{00}^{-1}S_{01}\frac{\partial}{\partial t_2}g(\boldsymbol{t}_0) + O(b^2)\\
\approx\ & g(\boldsymbol{t}_0) + \kappa^{-1}\kappa_{1^1}\frac{\partial}{\partial t_1}g(\boldsymbol{t}_0) + \kappa^{-1}\kappa_{2^1}\frac{\partial}{\partial t_1}g(\boldsymbol{t}_0) + O(b^2)\\
=\ & g(\boldsymbol{t}_0) + O(b)
\end{aligned}
$$

where the last equality holds since κ_{1^1} and κ_{2^1} are not 0 when $\boldsymbol{t}_0$ is close to the boundary. Recall from chapter 3 that the bias of the local linear estimator is still $O(b^2)$ for points close to the boundary, so we see an improvement in bias of order b when using the local linear estimator vs. the N-W estimator.

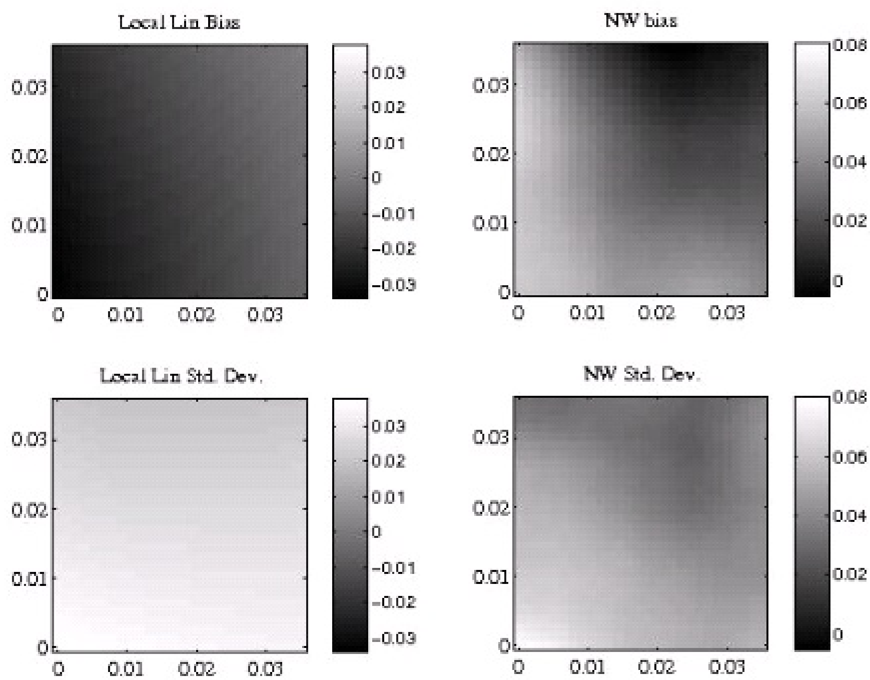

Figure 4.1: Mean and s.d. of the local linear and N-W estimators close to the boundary.

To examine this, we generated fine Gaussian random fields with $n = 1500$ and

$C(\boldsymbol{t}, \boldsymbol{t}+\boldsymbol{h}) = .4 - |\boldsymbol{h}|^{1.2} + .6|\boldsymbol{h}|^2$ in Matlab from code used to simulate the fields generated in Anderes and Chatterjee (2009). Then we introduce bias by multiplying a scaling field $D(\boldsymbol{t}) = 1+3t_1+t_2+.5t_1^2+.5t_2^2$. Figure 4.1 compares the NW estimator and our local linear estimator in estimating $f_{\boldsymbol{t}}\big((1,0)'\big)$ at the lower left corner of the unit square, using their optimal bandwidths respectively. The optimal rate for the N.W. estimator is $n^{-1/2}$ while the optimal rate for the local linear estimator is $n^{-2/3}$. Figure 4.1 shows that the bias and s.d. in estimating f at the boundary are smaller for the local linear estimator than the Nadaraya-Watson estimator.

4.2 Convergence Rates

Recall from Theorem 3.5

$$\begin{aligned}\hat{\beta}_0(\boldsymbol{t}_0; n, b) &\stackrel{d}{=} g(\boldsymbol{t}_0) + \frac{ZA(\boldsymbol{t}_0, b)^{1/2}}{(nb)^{d/2}} + \frac{1}{2}b^2[\mathcal{K}^{-1}\sum_{i,j=1}^{d}\frac{\partial}{\partial t_i}\frac{\partial}{\partial t_j}g(\boldsymbol{t}_0)\mathcal{N}_{i,j}]_1 \\ &\quad + o(b^2) + O\left(n^{-(\rho(\boldsymbol{t}_0)\wedge 2\delta(\boldsymbol{t}_0))}\right) + O_p\left(n^{-\delta(\boldsymbol{t}_0)}\sqrt{r_{nb}(\boldsymbol{t})}\right).\end{aligned}$$

First note that if the underlying process Y has a constant mean, or if $\mu(\boldsymbol{t})$ is smooth enough, then the terms with a $\delta(\boldsymbol{t}_0)$ are 0 or negligible. Now if we consider the terms that contain b, the best rate can be attained by equating the bias and s.d.

$$(nb)^{-d/2} = b^2 \Rightarrow n^{-d/2} = b^{2+d/2} \Rightarrow b = n^{-d/(4+d)}.$$

Therefore the convergence rate is $\max(n^{-2d/(4+d)}, n^{-\rho(\boldsymbol{t}_0)})$ which is optimal for estimating a regression function nonparametrically assuming two derivatives and iid. errors when $\rho(\boldsymbol{t}_0) > 2d/(4+d)$.

Recall from Theorem 3.8,

$$\begin{aligned}&\sup_{\boldsymbol{t}_0\in[0,1]^2}\left|\hat{\beta}_0(\boldsymbol{t}_0; n, b) - g(\boldsymbol{t}_0)\right| \\ = \;& O\left((nb)^{-d/2}\log(n)s_{nb}^{1/2}\right) + O\left((nb)^{-d/2}\log(n)r_{nb}^{1/2}n^{-\delta}\right) + O(b^2) + O(n^{-\rho}) + O(n^{-2\delta}).\end{aligned}$$

Again, if $\mu(\boldsymbol{t})$ is constant or smooth enough, any term with a δ is negligible. Then recall that s_{nb} depends on the value of ψ in relation to d, namely

$$s_{nb}(\boldsymbol{t}) = \begin{cases} 1 & \text{if } \psi(\boldsymbol{t}) > d/2 \\ \log(nb) & \text{if } \psi(\boldsymbol{t}) = d/2 \\ (nb)^{d-2\psi(\boldsymbol{t}_0)} & \text{if } \psi(\boldsymbol{t}) < d/2. \end{cases}$$

From Assumption 2.3(ii), we see that in practice $\psi(\boldsymbol{t}_0)$ can be made arbitrarily large as the order of the difference operator increases, so if we assume $\psi > d/2$.

$$(nb)^{-d/2}\log(n) = b^2 \Rightarrow n^{-d/2}\log(n) = b^{2+d/2} \Rightarrow b = n^{-d/(4+d)}(\log(n))^{2/(4+d)}.$$

Therefore, if we let $N = n^d$, the sample size, the convergence rate is

$$\max(N^{-2/(4+d)}\log(N)^{4/(4+d)}, N^{-\rho(\boldsymbol{t}_0)/d}).$$

Stone's optimal rate assuming two derivatives is $N^{-2/(4+d)}\log(N)^{2/(4+d)}$. So if we consider ρ sufficiently large, our rate is slightly worse than Stone's rate.

4.3 Separating $\alpha(\boldsymbol{t})$ and $f_{\boldsymbol{t}}(\boldsymbol{h})$

As in Chan and Wood (2000), we can consider estimating $\alpha(\boldsymbol{t})$.

Let

$$\xi(\boldsymbol{t},\boldsymbol{h};n,b) = n^{-\alpha(\boldsymbol{t})}\widehat{\beta}_0(\boldsymbol{t},\boldsymbol{h};n,b),$$

which does not depend on any unknown parameters. Define

$$\widehat{\alpha}(\boldsymbol{t};n,b) = \frac{\log\left(\xi(\boldsymbol{t},\boldsymbol{h};n,b_n)\right) - \log\left(\xi(\boldsymbol{t},\boldsymbol{h};2n,b_{2n})\right)}{\log 2}.$$

It follows that

$$\begin{aligned}
&\widehat{\alpha}(\boldsymbol{t};n,b) - \alpha(\boldsymbol{t}) \\
&= \frac{\log\left(\widehat{\beta}_0(\boldsymbol{t},\boldsymbol{h};n,b_n)\right) - \log\left(\widehat{\beta}_0(\boldsymbol{t},\boldsymbol{h};2n,b_{2n})\right)}{\log 2} \\
&= \frac{\log\left(1 + \frac{\widehat{\beta}_0(\boldsymbol{t},\boldsymbol{h};n,b_n) - J(\boldsymbol{0},\boldsymbol{t},\boldsymbol{h})}{J(\boldsymbol{0},\boldsymbol{t},\boldsymbol{h})}\right) - \log\left(1 + \frac{\widehat{\beta}_0(\boldsymbol{t},\boldsymbol{h};2n,b_{2n}) - J(\boldsymbol{0},\boldsymbol{t},\boldsymbol{h})}{J(\boldsymbol{0},\boldsymbol{t},\boldsymbol{h})}\right)}{\log 2} \\
&= O\left(\widehat{\beta}_0(\boldsymbol{t},\boldsymbol{h};n,b_n) - J(\boldsymbol{0},\boldsymbol{t},\boldsymbol{h}) + \widehat{\beta}_0(\boldsymbol{t},\boldsymbol{h};2n,b_{2n}) - J(\boldsymbol{0},\boldsymbol{t},\boldsymbol{h})\right)
\end{aligned}$$

Thus, $\widehat{\alpha}(\boldsymbol{t};n,b) - \alpha(\boldsymbol{t})$ has the same rate as $\widehat{\beta}_0(\boldsymbol{t},\boldsymbol{h};n,b_n) - J(\boldsymbol{0},\boldsymbol{t},\boldsymbol{h})$.

Now, let $\breve{\beta}_0(\boldsymbol{t},\boldsymbol{h};n,b)$ be the estimator defined with $\alpha(\boldsymbol{t})$ replaced by $\widehat{\alpha}(\boldsymbol{t};n,b)$ in $\widehat{\beta}_0(\boldsymbol{t},\boldsymbol{h};n,b)$; i.e.,

$$\breve{\beta}_0(\boldsymbol{t},\boldsymbol{h};n,b) = n^{\alpha(\boldsymbol{t})-\widehat{\alpha}(\boldsymbol{t};n,b)}\widehat{\beta}_0(\boldsymbol{t},\boldsymbol{h};n,b).$$

If we choose b so that $\widehat{\beta}_0(\boldsymbol{t},\boldsymbol{h};n,b_n) - J(\mathbf{0},\boldsymbol{t},\boldsymbol{h}) = O_p((\log n)^{-1})$, then

$$
\begin{aligned}
&\check{\beta}_0(\boldsymbol{t},\boldsymbol{h};n,b) - J(\mathbf{0},\boldsymbol{t},\boldsymbol{h}) \\
&= \left(n^{\alpha(\boldsymbol{t})-\widehat{\alpha}(\boldsymbol{t};n,b)} - 1\right)\widehat{\beta}_0(\boldsymbol{t},\boldsymbol{h};n,b) + \widehat{\beta}_0(\boldsymbol{t},\boldsymbol{h};n,b) - J(\mathbf{0},\boldsymbol{t},\boldsymbol{h}) \\
&= O_p(\log n)\left\{\widehat{\beta}_0(\boldsymbol{t},\boldsymbol{h};n,b) - J(\mathbf{0},\boldsymbol{t},\boldsymbol{h})\right\}.
\end{aligned}
$$

Thus, having to estimate $\alpha(\boldsymbol{t})$ in estimating $J(\mathbf{0},\boldsymbol{t},\boldsymbol{h})$ incurs the cost of a multiplicative rate of $\log n$ over the rate when $\alpha(\boldsymbol{t})$ is known. However, the extra cost does not make the rate prohibitively worse since $\widehat{\beta}_0(\boldsymbol{t},\boldsymbol{h};n,b) - J(\mathbf{0},\boldsymbol{t},\boldsymbol{h})$ has polynomial rate if we choose b sensibly.

4.4 Higher Order Differencing

From the results above, we see that $\psi(\boldsymbol{t})$ needs to be greater than d to achieve optimal convergence rates. In chapter 3 this corresponds to $x > \frac{\alpha(\boldsymbol{t})+d}{2}$. One question we can ask is what is the impact of using a higher order differencing operator than is necessary. To compare estimators, consider estimating $f(\boldsymbol{h})$ with a difference operator of order x, i.e. $W_n(\boldsymbol{t}) = n^{\alpha(\boldsymbol{t}_0)/2}\Delta^x_{\boldsymbol{h}/n}Y(\boldsymbol{t})$ and

$$
\hat{f}(\boldsymbol{h}) = \frac{\hat{\beta}_0(\boldsymbol{t}_0;n,b)}{\sum_{i=1}^x\sum_{j=0}^x(-1)^{i+j}\binom{x}{i}\binom{x}{j}|i-j|^{\alpha(\boldsymbol{t})}}.
$$

The variance of $\hat{f}$ is approximately

$$
2\int_{I(\boldsymbol{t}_0,b)} \bar{k}(\boldsymbol{z})^2 d\boldsymbol{z}\cdot\sum_{\boldsymbol{j}\in\mathbb{Z}^d}\left(\frac{\sum_{i=0}^x\sum_{j=0}^x(-1)^{i+j}\binom{x}{i}\binom{x}{j}f_{\boldsymbol{t}_0}(\boldsymbol{j}+(i-j)\boldsymbol{h})|\boldsymbol{j}+(i-j)\boldsymbol{h}|^{\alpha(\boldsymbol{t}_0)}}{|\boldsymbol{h}|^{\alpha(\boldsymbol{t}_0)}\sum_{i=0}^x\sum_{j=0}^x(-1)^{i+j}\binom{x}{i}\binom{x}{j}|i-j|^{\alpha(\boldsymbol{t}_0)}}\right)^2,
$$

and notationally let

$$
S = \left(\frac{\sum_{i=0}^x\sum_{j=0}^x(-1)^{i+j}\binom{x}{i}\binom{x}{j}f_{\boldsymbol{t}_0}(\boldsymbol{j}+(i-j)\boldsymbol{h})|\boldsymbol{j}+(i-j)\boldsymbol{h}|^{\alpha(\boldsymbol{t}_0)}}{|\boldsymbol{h}|^{\alpha(\boldsymbol{t}_0)}\sum_{i=0}^x\sum_{j=0}^x(-1)^{i+j}\binom{x}{i}\binom{x}{j}|i-j|^{\alpha(\boldsymbol{t}_0)}}\right)^2.
$$

Figure 4.2 gives the values of S for orders $x = 1,2,3,4,5$ for varying α when $d = 2$, $\boldsymbol{h} = (1,0)$ and $f_{\boldsymbol{t}}(\boldsymbol{h}) = 1$. We see that a given order has the lowest variance over a range of α and the optimal order is increasing in α.

The local behavior in $\mu(\boldsymbol{t})$ and the bias will also play a role in which x is optimal. In general though, the optimal x will be the smallest integer that satisfies the assumptions of chapter 3.

α	.25	.75	1.25	1.75	2.25	2.75	3.25	3.75	4.25	4.75
$x = 1$	1.35	1.67								
2	1.84	1.67	1.59	1.67	2.15	4.65				
3	2.22	2.08	1.96	1.90	1.89	1.98	2.21	2.69	3.86	8.90
4	2.55	2.43	2.32	2.25	2.20	2.20	2.24	2.34	2.52	2.82
5	2.85	2.74	2.64	2.58	2.53	2.50	2.50	2.53	2.60	2.71

Figure 4.2: Calculated values of S for varying x and α.

4.5 Missing Data

Here we will give an informal treatment of what happens when data is missing at random.

Suppose that observations of the process Y on our sampling grid are missing at random with probability p. Then let $X(\boldsymbol{t})$ be a bernoulli variable that is 0 if one of the points in $\Delta_{\boldsymbol{h}}^x$ is missing and 1 otherwise. Then $E(X(\boldsymbol{t})) = (1-p)^{x+1}$ since there are $x+1$ points in the operator $\Delta_{\boldsymbol{h}}^x$ and $E(X(\boldsymbol{t})X(\boldsymbol{s})) = (1-p)^{2(x+1)-y}$ where y is the number of points that are in both $\Delta_{\boldsymbol{h}}^x Y(\boldsymbol{t})$ and $\Delta_{\boldsymbol{h}}^x Y(\boldsymbol{s})$. So define

$$B(\boldsymbol{z}, \boldsymbol{h}) = \begin{cases} (1-p)^{x+1+y}, & \text{if } \boldsymbol{z} = \pm y\boldsymbol{h},\ y = 0, 1, \ldots, x \\ (1-p)^{2x+2}, & \text{otherwise} \end{cases}$$

and it follows that

$$\mathrm{Cov}(X(\boldsymbol{t}), X(\boldsymbol{s})) = \begin{cases} (1-p)^{x+1+y} - (1-p)^{2x+2}, & \text{if } \boldsymbol{z} = \pm y\boldsymbol{h},\ y = 0, 1, \ldots, x \\ 0, & \text{otherwise .} \end{cases}$$

Assuming that $X(\boldsymbol{t})$ and $Y(\boldsymbol{t})$ are independent of each other,

$$\begin{aligned} E\left\{X(\boldsymbol{t})(\Delta_{\boldsymbol{h}}^x Y(\boldsymbol{t}))^2\right\} &= (1-p)^{x+1} E(\Delta_{\boldsymbol{h}}^x Y(\boldsymbol{t}))^2, \\ E\left\{X(\boldsymbol{t})X(\boldsymbol{s})\Delta_{\boldsymbol{h}}^x Y(\boldsymbol{t})\Delta_{\boldsymbol{h}}^x Y(\boldsymbol{s})\right\} &= B(\boldsymbol{t}-\boldsymbol{s}, \boldsymbol{h}) E\left\{\Delta_{\boldsymbol{h}}^x Y(\boldsymbol{t})\Delta_{\boldsymbol{h}}^x Y(\boldsymbol{s})\right\}. \end{aligned}$$

It follows that $n^{\alpha(\boldsymbol{t})/2} X(\boldsymbol{t}) \Delta_{\boldsymbol{h}/n}^x Y(\boldsymbol{t})$ satisfies Assumption 3.1 and $\beta_0(\boldsymbol{t}_0; n, b)$ estimates $(1-p)^{x+1} g(\boldsymbol{t}_0)$. If we let $X'(\boldsymbol{t})$ denote the Bernoulli variable which is 0 if a point $\boldsymbol{t}$ is missing from our sampling grid and 1 if it is not missing with probability p. Then we can estimate p with

$$\hat{p} = n^{-d} \sum_{\boldsymbol{t}_i} X(\boldsymbol{t}_i).$$

$E(\hat{p}) = p$ and for m even,

$$\begin{aligned}
E(\hat{p}-p)^m &= n^{-md} \sum_{\boldsymbol{t}_{i_1}, \boldsymbol{t}_{i_2}, \ldots, \boldsymbol{t}_{i_m}} E\left\{(X(\boldsymbol{t}_1)-p)(X(\boldsymbol{t}_2)-p)\cdots(X(\boldsymbol{t}_m)-p)\right\} \\
&= n^{-md} \sum_{\boldsymbol{t}_i} E(X(\boldsymbol{t}_i)-p)^m \\
&= n^{-md+d} E(X(\boldsymbol{t}_1)-p)^m \\
&= n^{-(m-1)d} \left\{ \left[\sum_{i=0}^{m} (-1)^{i-1} \binom{m}{i} p^{i+1} \right] + (-1)^m p^m \right\} \\
&\leq n^{-(m-1)d}.
\end{aligned}$$

Then by Borel-Cantelli,

$$\begin{aligned}
\sum_{n=1}^{\infty} P\left(\frac{n^{d/2}}{\sqrt{\log(n)}} |\hat{p}-p| > c \right) &\leq \sum_{n=1}^{\infty} \frac{E e^{n^d(\hat{p}-p)^2}}{e^{c^2 \log(n)}} \\
&\leq \sum_{n=1}^{\infty} n^{-c^2} E \sum_{x=0}^{\infty} (n^d(\hat{p}-p)^2)^x / x! \\
&= \sum_{n=1}^{\infty} n^{-c^2} \sum_{x=0}^{\infty} n^{xd} E(\hat{p}-p)^{2x} / x! \\
&\leq \sum_{n=1}^{\infty} n^{-c^2} (2 + \sum_{x=2}^{\infty} n^{-(x-1)d} / x!) \\
&= \sum_{n=1}^{\infty} n^{-c^2} (2 + n^{-1} \sum_{x=2}^{\infty} n^{-(x-2)d} / x!) \\
&\leq \sum_{n=1}^{\infty} n^{-c^2} (2 + n^{-1} \sum_{x=2}^{\infty} 1/x!) \\
&\leq \sum_{n=1}^{\infty} n^{-c^2} (2 + e^1) \\
&< \infty,
\end{aligned}$$

when $c > 1$. So $\hat{p} \overset{a.s.}{\to} p$ with rate $\frac{\sqrt{\log(n)}}{n^{d/2}}$. And combining this with Theorem 3.8 and applying the continuous mapping theorem, $\frac{\hat{\beta}_0(\boldsymbol{t}; n, b)}{\hat{p}} \overset{a.s.}{\to} g(\boldsymbol{t})$ uniformly over $[0,1]^d$ with rate $\frac{\log(n)}{(nb)^{d/2}}$.

CHAPTER 5

Simulation Results

For numerical analysis, we focus on the two-dimensional Gaussian process with a nonstationary Matérn covariance and one-dimensional multifractional Brownian motion. For the Matérn example, we evaluate the performance of the local linear smoother in estimating f when ν is known. The estimator is compared over a bandwidth range and for different increment orders with theoretical justification. Then we analyze linear prediction by estimating the process at an unobserved spatial location which is $(\frac{7}{20n}, \frac{7}{20n})$ off of the grid. For comparison, we predict with Kriging using the true variogram, Kriging using an estimated variogram, Kriging with just an estimate of ν and local estimators. Kriging using the true variogram always performs the best, but kriging with an estimated variogram is only slightly worse. For multifractional Brownian motion, we examine local estimation of the smoothness parameter $H(\boldsymbol{t})$.

We used Matlab to generate a local window of points for the Gaussian random field specified in example 5.1 with grid size n and window width $2b$. The "unobserved" point was also generated so we can calculate the bias and s.d. of our Kriging estimators. To generate the fields, we calculated $Y = \Sigma^{1/2}Z$, where Z is an vector of $(2nb)^2 + 1$ independent $N(0,1)$ realizations and Σ is the symmetric $(2nb)^2+1 \times (2nb)^2+1$ covariance matrix where $\Sigma_{ij} = \mathrm{Cov}(Y_i, Y_j)$. We approximated $\Sigma^{1/2}$ by using the Matlab singular value decomposition function (SVD). For the multifractional Brownian motion example we used a similar procedure which generates the grid of size $1/5000$ across $[0, 1]$.

5.1 Nonstationary Matérn Covariance

Example 5.1. *Gaussian random field with mean function* $\mu(\boldsymbol{t}) = 10 + \cos(2\pi t_1) + \sin(\pi t_2)$ *and nonstationary Matérn covariance function*

$$C(\boldsymbol{t},\boldsymbol{s}) = D(\boldsymbol{t})D(\boldsymbol{s})\frac{\sigma^2 2^{1-\nu}}{\Gamma(\nu)}(\sqrt{2\nu}|M(\boldsymbol{t}-\boldsymbol{s})|)^{\nu}K_{\nu}(\sqrt{2\nu}|M(\boldsymbol{t}-\boldsymbol{s})|)$$

with $\sigma^2 = 1, \rho = 1$, ν *varying,* $M = \begin{bmatrix} .5 & .5 \\ -1 & 3 \end{bmatrix}$ *and* $D(\boldsymbol{t}) = .5 + t_1 + .8t_2 - t_1 t_2$.

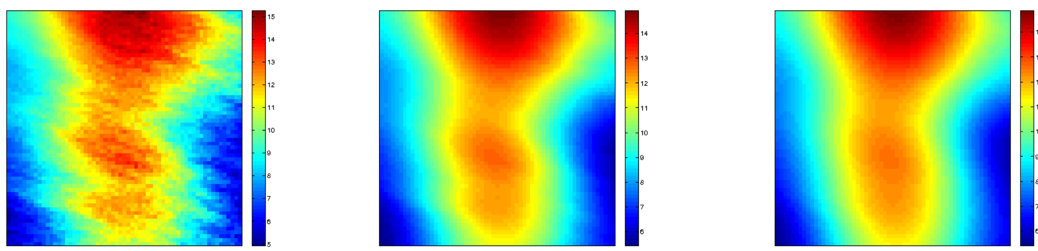

Figure 5.1: Realizations of Matérn random fields with $\nu = .4$(top), $\nu = 1.4$(left) and $\nu = 2.4$(right)

From figure 5.1 we see that as ν increases, the field becomes smoother. We can also see the anisotropy created by M.

Since ν is constant, we can accurately estimate ν globally if we have a large window of points. For this example we are only generating a local window of points, so we will assume that ν is known or accurately estimated.

5.1.1 Local Expansion

Here we will calculate the local expansion of the covariance function. Let $a = \sigma^2 \frac{2\pi}{\Gamma(\nu)\sin(\nu\pi)}$.

$$\begin{aligned} C(\boldsymbol{t},\boldsymbol{s}) &= aD(\boldsymbol{t})D(\boldsymbol{s})\sum_{m=0}^{\infty}\left\{\frac{(\nu/2)^m|M(\boldsymbol{t}-\boldsymbol{s})|^{2m}}{m!\Gamma(m-\nu+1)} - \frac{(\nu/2)^{m+\nu}|M(\boldsymbol{t}-\boldsymbol{s})|^{2m+2\nu}}{m!\Gamma(m+\nu+1)}\right\} \\ &= aD(\boldsymbol{t})D(\boldsymbol{s})\left\{\sum_{m=0}^{\lceil\nu\rceil}\frac{(\nu/2)^m\left(\sum_{j=1}^d\left(\sum_{k=1}^d M_{jk}(t_k-s_k)\right)^2\right)^m}{m!\Gamma(m-\nu+1)}\right. \\ &\qquad \left. -\frac{(\nu/2)^{\nu}|M(\boldsymbol{t}-\boldsymbol{s})|^{2\nu}}{\Gamma(\nu+1)}\right\} + O\left(|\boldsymbol{t}-\boldsymbol{s}|^{2\nu+1}\right) \end{aligned}$$

$$= \quad aD(\boldsymbol{t})D(\boldsymbol{s})\left\{\sum_{m=0}^{\lceil\nu\rceil}\sum_{|\boldsymbol{\ell}_1|+|\boldsymbol{\ell}_2|=2m} c_{\boldsymbol{\ell}_1,\boldsymbol{\ell}_2}\boldsymbol{t}^{\boldsymbol{\ell}_1}\boldsymbol{s}^{\boldsymbol{\ell}_2} + \frac{(\nu/2)^\nu|M(\boldsymbol{t}-\boldsymbol{s})|^{2\nu}}{\Gamma(\nu+1)}\right\} + O\left(|\boldsymbol{t}-\boldsymbol{s}|^{2\nu+1}\right),$$

for some constants $c_{\boldsymbol{\ell}_1,\boldsymbol{\ell}_2}$. We can express this as

$$\sum_{m=0}^{\lceil\nu\rceil}\sum_{|\boldsymbol{\ell}_1|+|\boldsymbol{\ell}_2|=2m} c_{\boldsymbol{\ell}_1,\boldsymbol{\ell}_2}\boldsymbol{t}^{\boldsymbol{\ell}_1}\boldsymbol{s}^{\boldsymbol{\ell}_2} = \sum_{|\boldsymbol{\ell}|=0}^{\lceil\nu\rceil} b_{\boldsymbol{\ell}}(\boldsymbol{t})\boldsymbol{s}^{\boldsymbol{\ell}} + \sum_{|\boldsymbol{\ell}|=0}^{\lceil\nu\rceil} b_{\boldsymbol{\ell}}(\boldsymbol{s})\boldsymbol{t}^{\boldsymbol{\ell}}$$

where $b_{\boldsymbol{\ell}}(\boldsymbol{t})$ are polynomials. Then by definition of D,

$$\begin{aligned}
& aD(\boldsymbol{t})D(\boldsymbol{s})\left(\sum_{|\boldsymbol{\ell}|=0}^{\lceil\nu\rceil} b_{\boldsymbol{\ell}}(\boldsymbol{t})\boldsymbol{s}^{\boldsymbol{\ell}} + \sum_{|\boldsymbol{\ell}|=0}^{\lceil\nu\rceil} b_{\boldsymbol{\ell}}(\boldsymbol{s})\boldsymbol{t}^{\boldsymbol{\ell}}\right) \\
= \quad & a(.5+t_1+.8t_2-t_1t_2)(.5+s_1+.8s_2-s_1s_2)\left(\sum_{|\boldsymbol{\ell}|=0}^{\lceil\nu\rceil} b_{\boldsymbol{\ell}}(\boldsymbol{t})\boldsymbol{s}^{\boldsymbol{\ell}} + \sum_{|\boldsymbol{\ell}|=0}^{\lceil\nu\rceil} b_{\boldsymbol{\ell}}(\boldsymbol{s})\boldsymbol{t}^{\boldsymbol{\ell}}\right) \\
= \quad & \sum_{|\boldsymbol{\ell}|=0}^{\lceil\nu+2\rceil} b_{\boldsymbol{\ell}}(\boldsymbol{t})\boldsymbol{s}^{\boldsymbol{\ell}} + \sum_{|\boldsymbol{\ell}|=0}^{\lceil\nu+2\rceil} b_{\boldsymbol{\ell}}(\boldsymbol{s})\boldsymbol{t}^{\boldsymbol{\ell}},
\end{aligned}$$

where $b_{\boldsymbol{\ell}}$ are redefined to account for D. Lastly,

$$aD(\boldsymbol{t})D(\boldsymbol{s})\frac{(\nu/2)^\nu|M(\boldsymbol{t}-\boldsymbol{s})|^{2\nu}}{\Gamma(\nu+1)} = aD(\boldsymbol{t})^2\frac{(\nu/2)^\nu|M(\boldsymbol{t}-\boldsymbol{s})|^{2\nu}}{\Gamma(\nu+1)} + O\left(|\boldsymbol{t}-\boldsymbol{s}|^{2\nu+1}\right).$$

Therefore this satisfies our covariance assumption with

$$f_{\boldsymbol{t}}(\boldsymbol{t}-\boldsymbol{s}) = \frac{\sigma^2 2\pi D(\boldsymbol{t})^2(\nu/2)^\nu}{\Gamma(\nu)\sin(\pi\nu)\Gamma(\nu+1)}|M(\frac{\boldsymbol{t}-\boldsymbol{s}}{|\boldsymbol{t}-\boldsymbol{s}|})|^{2\nu},$$

$\alpha(\boldsymbol{t}) = 2\nu$ and $\gamma(\boldsymbol{t}) = 1$.

5.1.2 Optimal Bandwidth, Increment Order and Normality

From section 4.2, the optimal bandwith will be $b = cn^{-1/3}$ for some constant c. To determine the optimal theoretical c, we will calculate the theoretical bias and variance of the local linear smoother. In these calculations recall that $M = \begin{bmatrix} .5 & .5 \\ -1 & 3 \end{bmatrix}$ and $D(\boldsymbol{s}) = .5 + s_1 + .8s_2 - s_1 * s_2$. Here we will set $\nu = .4$ ($\alpha = .8$), $x = 2$ and $\boldsymbol{h} = (1,0)'$ which gives $f_{\boldsymbol{t}_0}(\boldsymbol{h}) = .964$.

From Theorem 2.5, the bias is approximately

$$\frac{\frac{\partial^2}{\partial t_1^2} g(\boldsymbol{t}_0) \int k(\boldsymbol{z}) z_1^2 d\boldsymbol{z} + \frac{\partial^2}{\partial t_2^2} g(\boldsymbol{t}_0) \int k(\boldsymbol{z}) z_2^2 d\boldsymbol{z}}{2 \int k(\boldsymbol{z}) d\boldsymbol{z}}$$

So if we consider the local linear estimator with kernel function $k(\boldsymbol{z}) = (1 - z_1^2) * (1 - z_2^2)$, the theoretical bias is approximately

$$(32/9)^{-1}(16/45)\left(\frac{\partial^2}{\partial t_1^2} g(\boldsymbol{t}_0) + \frac{\partial^2}{\partial t_2^2} g(\boldsymbol{t}_0)\right) b^2 \approx .39b^2.$$

And from Theorem 2.7, the variance is $n^{-2}b^{-2}A(\boldsymbol{t}_0)$ where $A(\boldsymbol{t}_0)$ is

$$2\left(\int k(\boldsymbol{z}) d\boldsymbol{z}\right)^{-2} \int k(\boldsymbol{z})^2 d\boldsymbol{z} \sum_{\boldsymbol{j}} g^2(\boldsymbol{t}_0, \boldsymbol{j}) \approx 1.81.$$

So to minimize mean squared error,

$$\min_b(.39^2 b^4 + 1.81 n^{-2} b^{-2}) = \min_c(.39^2 c^4 + 1.81 c^{-2}) n^{-2/3}$$

which resuls in $c = 1.35$.

Figure 5.2 gives the results from 10,000 independent realizations of the field and estimation of $f_{\boldsymbol{t}}$ as c increases. First note that these values are consistent with our theoretical calculations. The figure confirms that the approximate minimum of the MSE plot is consistent with the 1.35 value.

Figure 5.3 gives the RMSE of the local linear smoother for varying ν and x when nb is fixed at 30 and c is fixed at 1.35. From section 4.4, we expect that the optimal x will increase as ν increases. For this example, $x = 2$ is optimal when $\nu < .5$, $x = 4$ is optimal when $\nu \in (.5, 2)$ and $x = 6$ is optimal when $\nu \in (2, 3.5)$. For larger values of ν, it appears that approximation errors dominate the error rates.

To evaluate asymptotic normality, figure 5.4 gives a comparison of Q-Q plots as nb increases. As expected, the estimators look more normal as nb increases.

5.1.3 Prediction

Now we will examine the error in predicting the value of $Y(\boldsymbol{t}_0)$ at an unobserved location $\boldsymbol{t}_0$ that is $(\frac{7}{20n}, \frac{7}{20n})$ off of the grid. The first estimator we consider is the nearest neighbor estimator defined as $\hat{Y}_{NN} := Y(\boldsymbol{t}_i)$ where $\boldsymbol{t}_i$ is the closest gridpoint

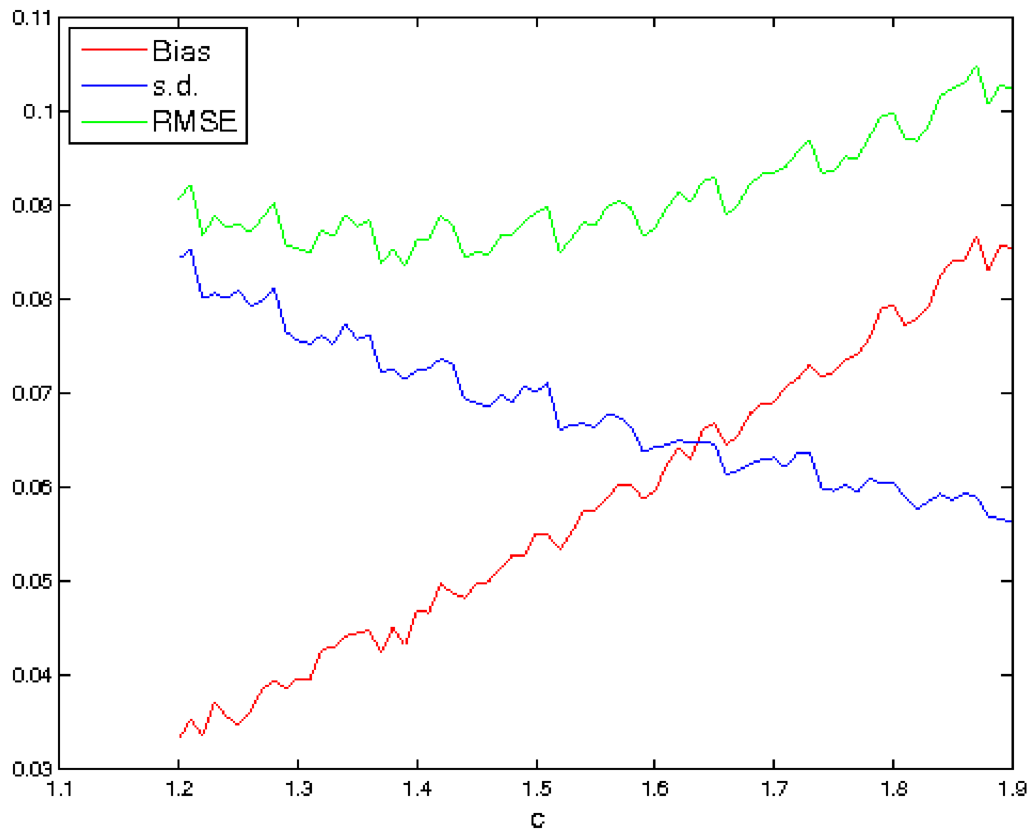

Figure 5.2: Bias, s.d. and RMSE for estimation of f and varying c.

to $\boldsymbol{t}_0$. Next is the universal Kriging estimator defined as $\hat{Y}_{UK-k} := \boldsymbol{\lambda}'\boldsymbol{Y}$ where

$$
\begin{aligned}
\boldsymbol{\lambda}' &= \left(\boldsymbol{v}_1 + \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{K}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{v}_2 - \boldsymbol{X}\boldsymbol{K}^{-1}\boldsymbol{v}_1)\right)' \boldsymbol{K}^{-1} \\
\boldsymbol{Y} &= (Y(\boldsymbol{t}_1), \ldots, Y(\boldsymbol{t}_m))' \\
\boldsymbol{\lambda} &= (\lambda_1, \ldots, \lambda_m)', \\
\boldsymbol{K} &= \{f(\boldsymbol{t}_i - \boldsymbol{t}_j)|\boldsymbol{t}_i - \boldsymbol{t}_j|^{2\nu}\}_{i,j=1}^m, \\
\boldsymbol{X} &= \text{ a matrix with } m \text{ rows where the } i\text{-th row is the vector containing} \\
& \quad\ \boldsymbol{t}_i^{\boldsymbol{\ell}}\text{; the order is arranged in any order, } \forall \boldsymbol{\ell} : |\boldsymbol{\ell}| \le r, \\
\boldsymbol{v}_1 &= (f(\boldsymbol{t}_1 - \boldsymbol{t}_0)|\boldsymbol{t}_1 - \boldsymbol{t}_0|^{2\nu}, \ldots, f(\boldsymbol{t}_m - \boldsymbol{t}_0)|\boldsymbol{t}_m - \boldsymbol{t}_0|^{2\nu})', \\
\boldsymbol{v}_2 &= \text{ the vetor containing } \boldsymbol{t}_0^{\boldsymbol{\ell}} \text{ arranged in the same order as } \boldsymbol{X}.
\end{aligned}
$$

For this estimator we will approximate $f_{\boldsymbol{t}}(\boldsymbol{u})$ with $|\hat{M}\boldsymbol{u}|^{2\nu}$, which is generated by estimating M from $\hat{f}_{\boldsymbol{t}_0}(\boldsymbol{h})$ for $\boldsymbol{h} \in \{(1,0),(0,1),(1,1)\}$ as in Anderes (2011). We

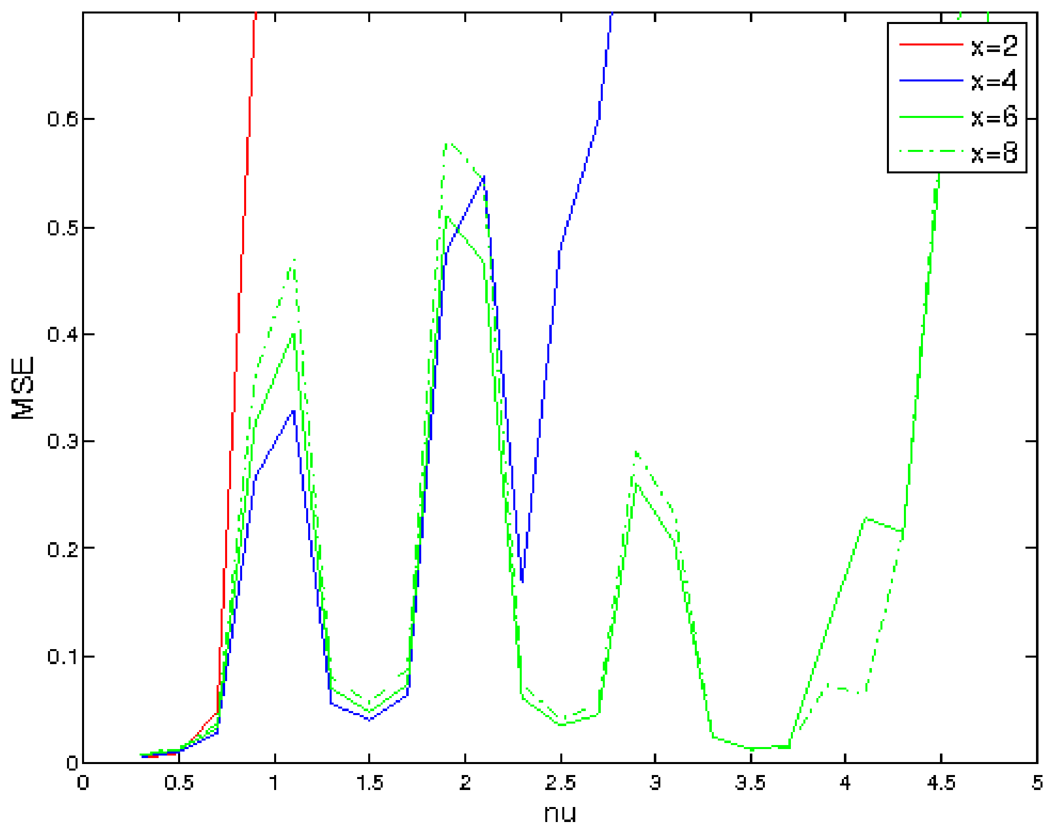

Figure 5.3: MSE for estimation of f for varying ν and x.

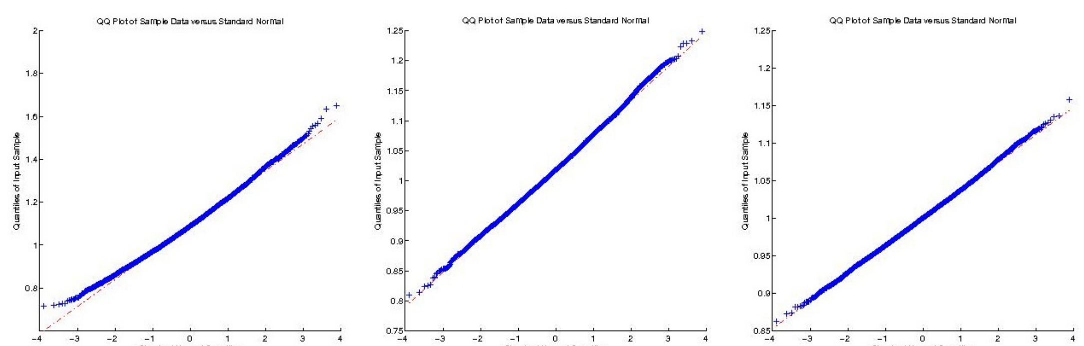

Figure 5.4: Q-Q plots of $\hat{f}'$ for $nb = 10$(left), $nb = 15$(center) and $nb = 20$(right).

also consider the universal Kriging estimator assuming isotropy and using the true ν, labeled $\hat{Y}_{IK-k}$, and the linear predictor whose coefficients eliminate polynomials of degree k, labeled $\hat{Y}_{LP-k}$. Lastly, we will evaluate the universal Kriging estimator using the true variogram. This will be labeled as $TK - k$, where $\hat{Y}_{TK-k}$ is defined the same as $\hat{Y}_{UK-k}$ except the estimated principal irregular term is replaced with the true variogram. The $TK - k$ estimators will have the lowest prediction s.d.'s since they are using the true variogram, but we expect that the $UK - k$ estimators will perform similarly if M is accurately estimated. For consistency, each estimator will predict with the 12×12 grid of points centered at $\boldsymbol{t}_0$. We fix $nb = 30$ and for the $\hat{Y}_{UK-k}$ estimators, we will use $x = 4$ when $\nu < 2$ and $x = 8$ when $\nu > 2$.

Figure 5.5 gives the prediction s.d. for the different estimators as ν increases. For simplicity of presentation, we multiplied each value by 1000. The $\hat{Y}_{TK-2}$ and $\hat{Y}_{TK-4}$ estimators performed the same as $\hat{Y}_{TK-0}$, so they were left out. We also only include the $\hat{Y}_{LP-2}$ and $\hat{Y}_{IK-4}$ estimators. As expected, $\hat{Y}_{TK-0}$ has the lowest s.d. across ν and $\hat{Y}_{NN}$ has the largest s.d. across ν. $\hat{Y}_{UK-4}$ performed the best among the $\hat{Y}_{UK-k}$ estimators, the s.d. of $\hat{Y}_{TK-0}$ is about 5% lower than the s.d. of $\hat{Y}_{UK-4}$ when $\nu = .3$ and there is less than a 1% difference when $\nu \geq .7$. The $LP-2$ and $IK-4$ estimators perform significantly worse than the $UK-4$ estimator, suggesting that it is beneficial to estimate the variogram.

From figure 5.6 we see that the bias is small compared to the s.d. for all of the estimators except $\hat{Y}_{NN}$, which has a larger bias when $\nu > 1.3$.

ν	.3	.7	1.3	1.7	2.3	2.7
s.d.($\hat{Y}_{NN}$)	294.2	55.9	16.97	13.15	11.6	11.4
s.d.($\hat{Y}_{LP-2}$)	282.3	45.6	4.27	1.03	.163	.059
s.d.($\hat{Y}_{IK-4}$)	248.3	44.9	4.31	.97	.116	.029
s.d.($\hat{Y}_{UK-0}$)	236.7	35.4	2.34	.416	.0369	.0077
s.d.($\hat{Y}_{UK-2}$)	248.3	34.1	2.34	.416	.0365	..0074
s.d.($\hat{Y}_{UK-4}$)	230.4	33.9	2.34	.416	.0363	.0073
s.d.($\hat{Y}_{TK-0}$)	223.2	33.6	2.33	.414	.0363	.0072

Figure 5.5: Prediction standard deviation for different values of ν, multiplied by 1000.

5.2 Multifractional Brownian Motion

This simulation examines local estimation of the Hurst parameter in one-dimensional mBm.

From Ayache (2000), the covariance for one-dimensional mBm can be written as

$$C(t, s) = D(H(s), H(t)) \left(t^{H(t)+H(s)} + s^{H(t)+H(s)} - |t - s|^{H(t)+H(s)} \right),$$

ν	.3	.7	1.3	1.7	2.3	2.7
$\overline{\hat{Y}}_{NN} - \overline{Y}$	-4.6	-26.7	-24.3	-22.1	-21.7	-21.7
$\overline{\hat{Y}}_{LP-2} - \overline{Y}$	15.4	$-.3$	$-.11$	$-.016$	$-.004$	.002
$\overline{\hat{Y}}_{IK-4} - \overline{Y}$	4.1	.21	$-.19$	$-.03$	$-.002$	.003
$\overline{\hat{Y}}_{UK-0} - \overline{Y}$	18.5	$-.43$	.163	.014	$-.013$	$-.0028$
$\overline{\hat{Y}}_{UK-2} - \overline{Y}$	4.6	$-.68$	.09	$-.023$	$-.005$	$-.0018$
$\overline{\hat{Y}}_{UK-4} - \overline{Y}$	3.2	$-.71$	.09	.013	.001	.0001
$\overline{\hat{Y}}_{TK-0} - \overline{Y}$	1.93	-1.02	.234	.081	.001	.0002

Figure 5.6: Prediction bias for different values of ν, multiplied by 1000.

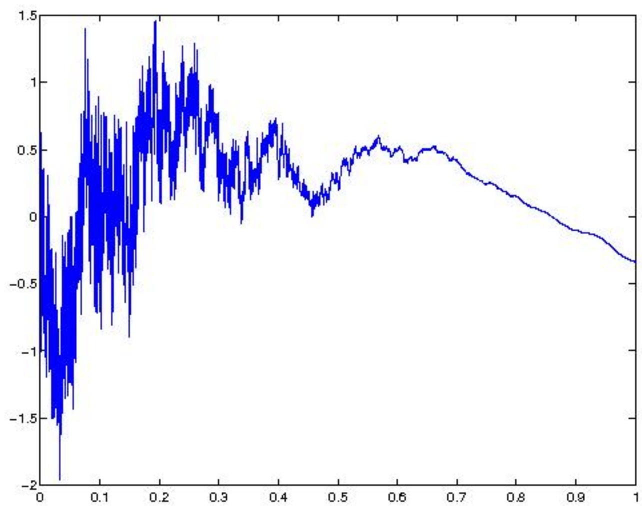

Figure 5.7: One realization of one dimensional mBm.

where

$$D(H(t), H(s)) = \frac{\sqrt{\Gamma(2H(t)+1)\Gamma(H(s)+1)\sin(\pi H(t))\sin(\pi H(s))}}{2\Gamma(H(t)+H(s)+1)\sin(\pi(H(t)+H(s))/2)},$$

and we let $t, s \in (0, 1]$. In this simulation we consider $H(t) = .5 - .4\cos(\pi t)$, which increases smoothly from .1 to .9 as t increases from 0 to 1. Figure 5.7 plots a single realization of this process, which becomes smoother as t (and $H(t)$) increases.

Now we will consider estimating $H(t)$ locally with a local linear smoother. Since this process is one-dimensional, our theoretical results tell us that the optimal rate is achieved when $b = cn^{-1/5}$ for some c. For this example, simulation determined that $c \approx .7$ is optimal. Figure 5.8 plots the true value of $H(t)$, the mean of the local linear smoother from 10,000 independent replications and the mean $\pm$ 2 standard deviations with $n = 5000$, $c = .7$ and $x = 4$.

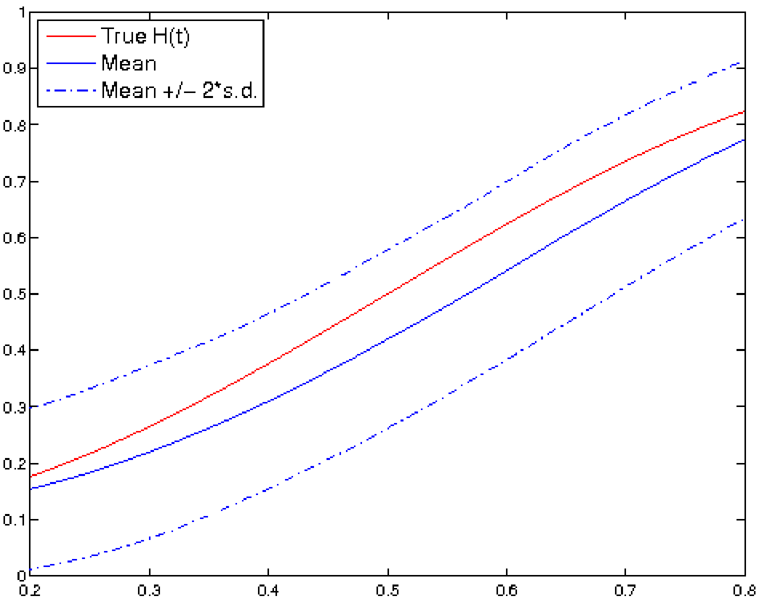

Figure 5.8: Estimation of $H(t)$ for $t \in [.2, .8]$.

APPENDICES

APPENDIX A

Proofs of Examples from Chapter 2

Multifractional Brownian Motion

Fractional Brownian motion(fBm) was introduced by Mandelbrot and Van Ness (1968) and defined for $t \geq 0$ and $H \in (0,1)$ as

$$B_H(t) = \frac{1}{\Gamma(H+1/2)}\left\{ \int_{-\infty}^{0} [(t-s)^{H-1/2} - (-s)^{H-1/2}]dW(s) + \int_{0}^{t} (t-s)^{H-1/2} dW(s) \right\},$$

where W is a Weiner process defined on $(-\infty, \infty)$. Herbin (2006) points out that we can also define fractional Brownian motion as a centered Gaussian process B_H s.t. $\forall \boldsymbol{s}, \boldsymbol{t} \in \mathbb{R}_+^d$,

$$E\left(B_H(\boldsymbol{s})B_H(\boldsymbol{t})\right) = \frac{1}{2}\left(\boldsymbol{s}^{2H} + \boldsymbol{t}^{2H} - |\boldsymbol{t}-\boldsymbol{s}|^{2H}\right).$$

This process has some desirable properties such as self-similarity, $B_H(a\boldsymbol{t}) \overset{d}{=} |a|B_H(\boldsymbol{t})$; stationary increments, $B_H(\boldsymbol{t}) - B_H(\boldsymbol{s}) \overset{d}{=} B_H(\boldsymbol{t}-\boldsymbol{s})$; long-range depenedence when $H > 1/2$; sample paths are nowhere differentiable with probability one.

Multifractional Brownian motion(mBm) is an extension of fBm where $H(\boldsymbol{t})$ is a Hölder function s.t. $0 < \alpha \leq H(\boldsymbol{t}) \leq \beta$ and $|H(\boldsymbol{t}) - H(\boldsymbol{s})| \leq c|\boldsymbol{t}-\boldsymbol{s}|^{\beta}$. Herbin (2006) showed that for a mBm process X,

$$E(X(\boldsymbol{s})X(\boldsymbol{t})) = D(H(\boldsymbol{s}), H(\boldsymbol{t}))\left\{|\boldsymbol{s}|^{H(\boldsymbol{s})+H(\boldsymbol{t})} + |\boldsymbol{t}|^{H(\boldsymbol{s})+H(\boldsymbol{t})} - |\boldsymbol{t}-\boldsymbol{s}|^{H(\boldsymbol{s})+H(\boldsymbol{t})}\right\},$$

where

$$D(H(\boldsymbol{t}), H(\boldsymbol{s})) = \int\limits_{\mathbb{R}^d} (1 - e^{iu_1})/(|bu|^{H(\boldsymbol{t})+H(\boldsymbol{s})+d}) d\boldsymbol{u}.$$

Proposition A.1. *Let H be three times differentiabe and assume we observe mBm away from $\mathbf{0}$. Then mBm satisfies our covariance assumption with $f_{\boldsymbol{t}}(\boldsymbol{s} - \boldsymbol{t}) = D(H(\boldsymbol{t}), H(\boldsymbol{t}))$, $\alpha(\boldsymbol{t}) = 2H(\boldsymbol{t})$, $r = 2$ and $\gamma(\boldsymbol{t}) < 1$.*

Proof. Since $D(\boldsymbol{t}, \boldsymbol{s})$ is a smooth function,

$$D^{(i,j)}(t, s) := \frac{\partial^i}{\partial t^i} \frac{\partial^j}{\partial s^j} D(t, s) \leq c_{i,j},$$

where $c_{i,j}$ are bounded for $i + j \leq 4$. Since $H(\boldsymbol{t})$ is three times differentiable for $|\boldsymbol{t}|$ bounded away from 0. Let $H^{(\boldsymbol{\ell})}(\boldsymbol{t}) = \frac{\partial^{\ell_1}}{\partial t_1^{\ell_1}} \cdots \frac{\partial^{\ell_d}}{\partial t_d^{\ell_d}} H(\boldsymbol{t})$. Then

$$|\boldsymbol{t}|^{H(\boldsymbol{s})+H(\boldsymbol{t})} = e^{(H(\boldsymbol{s})+H(\boldsymbol{t}))\log(|\boldsymbol{t}|)} = |\boldsymbol{t}|^{2H(\boldsymbol{t})} \left(\sum_{i=0}^{\infty} \left\{ \log(|\boldsymbol{t}|) \sum_{|\boldsymbol{\ell}|=1}^{2} (\boldsymbol{s} - \boldsymbol{t})^{\boldsymbol{\ell}} H^{(\boldsymbol{\ell})}(\boldsymbol{t}) \right\}^i \right).$$

Thus,

$$\begin{aligned}
& D(H(\boldsymbol{s}), H(\boldsymbol{t}))|\boldsymbol{t}|^{H(\boldsymbol{s})+H(\boldsymbol{t})} \\
= & \left(\sum_{|\boldsymbol{\ell}|=0}^{2} H^{(\boldsymbol{\ell})}(\boldsymbol{t}) D^{(|\boldsymbol{\ell}|,0)}(H(\boldsymbol{t}), H(\boldsymbol{t}))(\boldsymbol{s} - \boldsymbol{t})^{\boldsymbol{\ell}} + O(|\boldsymbol{s} - \boldsymbol{t}|^3) \right) \\
& \cdot \left(|\boldsymbol{t}|^{2H(\boldsymbol{t})} \sum_{i=0}^{2} \frac{1}{i!} \left\{ \log(|\boldsymbol{t}|) \left(\sum_{|\boldsymbol{\ell}|=1}^{2} H^{(\boldsymbol{\ell})}(\boldsymbol{t})(\boldsymbol{s} - \boldsymbol{t})^{\boldsymbol{\ell}} \right) \right\}^i + O(|\boldsymbol{s} - \boldsymbol{t}|^3) \right) \\
= & \sum_{|\boldsymbol{\ell}_1|+|\boldsymbol{\ell}_2|=0}^{2} H^{(\boldsymbol{\ell}_1)}(\boldsymbol{t}) H^{(\boldsymbol{\ell}_2)}(\boldsymbol{t}) D^{(|\boldsymbol{\ell}_1|,0)}(H(\boldsymbol{t}), H(\boldsymbol{t})) |\boldsymbol{t}|^{2H(\boldsymbol{t})} (\boldsymbol{s} - \boldsymbol{t})^{\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2} + O(|\boldsymbol{s} - \boldsymbol{t}|^3) \\
= & \sum_{|\boldsymbol{\ell}|=0}^{2} b_{\boldsymbol{\ell}}(\boldsymbol{t}) \boldsymbol{s}^{\boldsymbol{\ell}} + O(|\boldsymbol{t} - \boldsymbol{s}|^{\alpha(\boldsymbol{t})+\gamma(\boldsymbol{t})})
\end{aligned}$$

for bounded functions $b_{\boldsymbol{\ell}}(\boldsymbol{t})$, $\alpha(\boldsymbol{t}) = 2H(\boldsymbol{t})$ and $\gamma(\boldsymbol{t}) = 3 - \alpha(\boldsymbol{t})$. By symmetry, we also have

$$D(H(\boldsymbol{s}), H(\boldsymbol{t}))|\boldsymbol{s}|^{H(\boldsymbol{s})+H(\boldsymbol{t})} = \sum_{|\boldsymbol{\ell}|=0}^{2} b_{\boldsymbol{\ell}}(\boldsymbol{s}) \boldsymbol{t}^{\boldsymbol{\ell}} + O(|\boldsymbol{t} - \boldsymbol{s}|^{\alpha(\boldsymbol{t})+1}).$$

Then

$$
\begin{aligned}
& D(H(\boldsymbol{s}),H(\boldsymbol{t}))|\boldsymbol{s}-\boldsymbol{t}|^{H(\boldsymbol{s})+H(\boldsymbol{t})} \\
= & \left(\sum_{|\boldsymbol{\ell}|=0}^{2} H^{(\boldsymbol{\ell})}(\boldsymbol{t}) D^{(|\boldsymbol{\ell}|,0)}(H(\boldsymbol{t}),H(\boldsymbol{t}))(\boldsymbol{s}-\boldsymbol{t})^{\boldsymbol{\ell}} + O(|\boldsymbol{s}-\boldsymbol{t}|^3)\right) \\
& \cdot \left(|\boldsymbol{s}-\boldsymbol{t}|^{2H(\boldsymbol{t})}\left\{1+\log(|\boldsymbol{s}-\boldsymbol{t}|)\sum_{|\boldsymbol{\ell}|=1}^{2} H^{(\boldsymbol{\ell})}(\boldsymbol{t})(\boldsymbol{s}-\boldsymbol{t})^{\boldsymbol{\ell}}\right\} + O(|\boldsymbol{s}-\boldsymbol{t}|^{2H(\boldsymbol{t})+1})\right) \\
= & |\boldsymbol{s}-\boldsymbol{t}|^{2H(\boldsymbol{t})} D(H(\boldsymbol{t}),H(\boldsymbol{t})) + O\left(|\boldsymbol{s}-\boldsymbol{t}|^{2H(\boldsymbol{t})+1}\log(|\boldsymbol{s}-\boldsymbol{t}|)\right),
\end{aligned}
$$

which implies that $f_{\boldsymbol{t}}(\boldsymbol{s}-\boldsymbol{t}) = D(H(\boldsymbol{t}),H(\boldsymbol{t}))$, $\alpha(\boldsymbol{t}) = 2H(\boldsymbol{t})$ and $\gamma(\boldsymbol{t}) < 1$.

For the derivative bound, notice first that D is smooth, H is three times differentiable and $|\boldsymbol{t}|$ is bounded away from 0. Therefore $D(H(\boldsymbol{s}),H(\boldsymbol{t}))|\boldsymbol{t}|^{H(\boldsymbol{s})+H(\boldsymbol{t})}$ is three times continuously differentiable in $\boldsymbol{t}$ and $\boldsymbol{s}$ and hence $\partial_{\boldsymbol{u}}^{(1,1)}\partial_{\boldsymbol{v}}^{(1,1)} D(H(\boldsymbol{s}),H(\boldsymbol{t}))|\boldsymbol{t}|^{H(\boldsymbol{s})+H(\boldsymbol{t})}$ is uniformly bounded. Then since $\partial_{\boldsymbol{u}}^{(1,0)}\log(|\boldsymbol{t}-\boldsymbol{s}|) = |\boldsymbol{t}-\boldsymbol{s}|^{-2}\sum_{|\boldsymbol{\ell}|=1}(\boldsymbol{t}-\boldsymbol{s})^{\boldsymbol{\ell}}\boldsymbol{u}^{\boldsymbol{\ell}}$,

$$
\begin{aligned}
& \partial_{\boldsymbol{u}}^{(1,0)} D(H(\boldsymbol{t}),H(\boldsymbol{s}))|\boldsymbol{t}-\boldsymbol{s}|^{H(\boldsymbol{t})+H(\boldsymbol{s})} \\
= & D(H(\boldsymbol{t}),H(\boldsymbol{s}))|\boldsymbol{t}-\boldsymbol{s}|^{H(\boldsymbol{t})+H(\boldsymbol{s})} \\
& \qquad \cdot \left\{(H(\boldsymbol{t})+H(\boldsymbol{s}))|\boldsymbol{t}-\boldsymbol{s}|^{-2}\sum_{|\boldsymbol{\ell}|=1}(\boldsymbol{t}-\boldsymbol{s})^{\boldsymbol{\ell}}\boldsymbol{u}^{\boldsymbol{\ell}} + \log(|\boldsymbol{t}-\boldsymbol{s}|)\sum_{|\boldsymbol{\ell}|=1}\boldsymbol{u}^{\boldsymbol{\ell}} H^{(\boldsymbol{\ell})}(\boldsymbol{t})\right\} \\
& + |\boldsymbol{t}-\boldsymbol{s}|^{H(\boldsymbol{t})+H(\boldsymbol{s})} D^{(1,0)}(H(\boldsymbol{t}),H(\boldsymbol{s}))\sum_{|\boldsymbol{\ell}|=1}\boldsymbol{u}^{\boldsymbol{\ell}} H^{(\boldsymbol{\ell})}(\boldsymbol{t}) \\
= & O(|\boldsymbol{t}-\boldsymbol{s}|^{H(\boldsymbol{t})+H(\boldsymbol{s})-1}).
\end{aligned}
$$

And in general, derivatives will be a sum of terms with the form

$$
\begin{aligned}
& R(\boldsymbol{t},\boldsymbol{s},\boldsymbol{u},\boldsymbol{v})|\boldsymbol{t}-\boldsymbol{s}|^{H(\boldsymbol{t})+H(\boldsymbol{s})-x}\left(\sum_{|\boldsymbol{\ell}|=1}(\boldsymbol{t}-\boldsymbol{s})^{\boldsymbol{\ell}}\boldsymbol{u}^{\boldsymbol{\ell}}\right)^{i}\left(\sum_{|\boldsymbol{\ell}|=1}(\boldsymbol{t}-\boldsymbol{s})^{\boldsymbol{\ell}}\boldsymbol{v}^{\boldsymbol{\ell}}\right)^{j}\log^{r}(|\boldsymbol{t}-\boldsymbol{s}|) && \text{(A.1)} \\
= & O(|\boldsymbol{t}-\boldsymbol{s}|^{H(\boldsymbol{t})+H(\boldsymbol{s})-x+i+j-r}), && \text{(A.2)}
\end{aligned}
$$

where $R(\boldsymbol{t},\boldsymbol{s},\boldsymbol{u},\boldsymbol{v})$ is a smooth function in $\boldsymbol{t}$ and $\boldsymbol{s}$. Then if we take the directional derivative of one of the terms in (A.1), we will add at most a $|\boldsymbol{t}-\boldsymbol{s}|^{-1}$ to the $O(\cdot)$ term in (A.2),

$$
\begin{aligned}
\partial_{\boldsymbol{u}}^{(1,0)} R(\boldsymbol{t},\boldsymbol{s},\boldsymbol{u},\boldsymbol{v}) &= O(1) \\
\partial_{\boldsymbol{u}}^{(1,0)}|\boldsymbol{t}-\boldsymbol{s}|^{H(\boldsymbol{t})+H(\boldsymbol{s})-x} &= (H(\boldsymbol{t})+H(\boldsymbol{s})-x)|\boldsymbol{t}-\boldsymbol{s}|^{H(\boldsymbol{t})+H(\boldsymbol{s})-x-2}\sum_{|\boldsymbol{\ell}|=1}(\boldsymbol{t}-\boldsymbol{s})^{\boldsymbol{\ell}}\boldsymbol{u}^{\boldsymbol{\ell}}
\end{aligned}
$$

$$
\begin{aligned}
&= O(|\boldsymbol{t}-\boldsymbol{s}|^{H(\boldsymbol{t})+H(\boldsymbol{s})-x-1}) \\
\partial_{\boldsymbol{u}}^{(1,0)} \left(\sum_{|\boldsymbol{\ell}|=1} (\boldsymbol{t}-\boldsymbol{s})^{\boldsymbol{\ell}} \boldsymbol{u}^{\boldsymbol{\ell}} \right)^{i} &= i \left(\sum_{|\boldsymbol{\ell}|=1} (\boldsymbol{t}-\boldsymbol{s})^{\boldsymbol{\ell}} \boldsymbol{u}^{\boldsymbol{\ell}} \right)^{i-1} \sum_{|\boldsymbol{\ell}|=1} \boldsymbol{u}^{2\boldsymbol{\ell}} \\
&= O(|\boldsymbol{t}-\boldsymbol{s}|^{i-1}) \\
\partial_{\boldsymbol{u}}^{(1,0)} \left(\sum_{|\boldsymbol{\ell}|=1} (\boldsymbol{t}-\boldsymbol{s})^{\boldsymbol{\ell}} \boldsymbol{v}^{\boldsymbol{\ell}} \right)^{j} &= j \left(\sum_{|\boldsymbol{\ell}|=1} (\boldsymbol{t}-\boldsymbol{s})^{\boldsymbol{\ell}} \boldsymbol{v}^{\boldsymbol{\ell}} \right)^{j-1} \sum_{|\boldsymbol{\ell}|=1} \boldsymbol{u}^{\boldsymbol{\ell}} \boldsymbol{v}^{\boldsymbol{\ell}} \\
&= O(|\boldsymbol{t}-\boldsymbol{s}|^{j-1}).
\end{aligned}
$$

Thus,

$$
\partial_{\boldsymbol{v}}^{(1,1)} \partial_{\boldsymbol{u}}^{(1,1)} D(H(\boldsymbol{t}), H(\boldsymbol{s})) |\boldsymbol{t}-\boldsymbol{s}|^{H(\boldsymbol{t})+H(\boldsymbol{s})} = O(|\boldsymbol{t}-\boldsymbol{s}|^{H(\boldsymbol{t})+H(\boldsymbol{s})-4}),
$$

which satisfies our assumption. □

Anisotropic Matérn Covariance

The general form for the Matérn covariance is given by

$$
C(t) = \frac{\sigma^2 2^{1-\nu}}{\Gamma(\nu)} (\sqrt{2\nu}|t|)^{\nu} K_{\nu}(\sqrt{2\nu}|t|),
$$

where K_ν is the modified Bessel function of the second kind.

The modified Bessel function of the first kind is defined by the infinite series:

$$
I_\nu(x) = \sum_{m=0}^{\infty} \frac{1}{m!\Gamma(m+\nu+1)} \left(\frac{x}{2}\right)^{2m+\nu}.
$$

The modified Bessel function of the second kind is defined as

$$
K_\nu(x) = \frac{\pi}{2} \frac{I_{-\nu}(x) - I_\nu(x)}{\sin(\nu\pi)},
$$

for non-integer ν. For integer ν, K_ν is defined as a limit. Then for non-integer ν,

$$
\begin{aligned}
K_\nu(x) &= \frac{\pi}{2} (\sin(\nu\pi))^{-1} \\
&\quad \cdot \sum_{m=0}^{\infty} \left\{ \frac{1}{m!\Gamma(m-\nu+1)} \left(\frac{x}{2}\right)^{2m-\nu} - \frac{1}{m!\Gamma(m+\nu+1)} \left(\frac{x}{2}\right)^{2m+\nu} \right\}
\end{aligned}
$$

Then the Matern covariance function is

$$
\begin{aligned}
& R_\nu(|\boldsymbol{t}|) \\
= \; & \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\frac{|\boldsymbol{t}|}{\rho}\right)^\nu K_\nu\left(\sqrt{2\nu}\frac{|\boldsymbol{t}|}{\rho}\right) \\
= \; & \sigma^2 \frac{2\pi}{\Gamma(\nu)} \left(\frac{\sqrt{\nu}|\boldsymbol{t}|}{\sqrt{2}\rho}\right)^\nu (\sin(\nu\pi))^{-1} \\
& \cdot \sum_{m=0}^{\infty} \left\{ \frac{1}{m!\Gamma(m-\nu+1)} \left(\frac{\sqrt{\nu}|\boldsymbol{t}|}{\sqrt{2}\rho}\right)^{2m-\nu} - \frac{1}{m!\Gamma(m+\nu+1)} \left(\frac{\sqrt{\nu}|\boldsymbol{t}|}{\sqrt{2}\rho}\right)^{2m+\nu} \right\} \\
= \; & \sigma^2 \frac{2\pi}{\Gamma(\nu)\sin(\nu\pi)} \sum_{m=0}^{\infty} \left\{ \frac{(\nu/2)^m |\boldsymbol{t}|^{2m}}{m!\Gamma(m-\nu+1)\rho^{2m}} - \frac{(\nu/2)^{m+\nu}|\boldsymbol{t}|^{2m+2\nu}}{m!\Gamma(m+\nu+1)\rho^{2m+2\nu}} \right\}.
\end{aligned}
$$

Anderes (2010) considers the geometric anisotropic Matérn covariance function, which can be defined as the covariance of a Gaussian random field $Y(\boldsymbol{t}) = Z(M\boldsymbol{t})$, where M is an invertible matrix with determinant 1 and Z is an isotropic Gaussian random field with a Matérn covariance. When $\nu \in (0,1)$, this covariance can be written as

$$
\begin{aligned}
& R_\nu(|M\boldsymbol{t}|) \\
= \; & \sigma^2 \frac{2\pi}{\Gamma(\nu)\sin(\nu\pi)} \sum_{m=0}^{\infty} \left\{ \frac{(\nu/2)^m |M\boldsymbol{t}|^{2m}}{m!\Gamma(m-\nu+1)\rho^{2m}} - \frac{(\nu/2)^{m+\nu}|M\boldsymbol{t}|^{2m+2\nu}}{m!\Gamma(m+\nu+1)\rho^{2m+2\nu}} \right\},
\end{aligned}
$$

which satisfies assumption 2.3 (i) with $b_{\boldsymbol{0}} = \frac{\sigma^2\pi}{\Gamma(\nu)\sin(\nu\pi)\Gamma(1-\nu)}$, $f_{\boldsymbol{t}}(\boldsymbol{t}-\boldsymbol{s}) = \frac{\sigma^2\pi\nu^\nu |M\frac{\boldsymbol{t}-\boldsymbol{s}}{|\boldsymbol{t}-\boldsymbol{s}|}|^{2\nu}}{\Gamma(\nu)\sin(\nu\pi)\rho^{2\nu}\Gamma(1+\nu)}$, $\alpha(\boldsymbol{t}) = 2\nu$, $\gamma(\boldsymbol{t}) = 2$ and $b_{\boldsymbol{\ell}}(\boldsymbol{s})$ can clearly be defined from $|M(\boldsymbol{t}-\boldsymbol{s})|^{2|\boldsymbol{\ell}|}$.

Deformation Model

Anderes and Chatterjee (2009) consider estimation of a deformation function F in a framework similar to ours. They assume that the observed process has the form $Y(\boldsymbol{t}) = Z(F(\boldsymbol{t}))$ where Z is assumed to be an isotropic Gaussian random field. They assume the following assumptions on the spatial process Z and covariance function R:

- R1: Z is a constant mean Gaussian process on $\mathbb{R}^2$ with autocovariance $R(|t-s|) = \mathrm{Cov}(Z(t), Z(s))$.
- R2: $R(|t|) = R(0) - |t|^\alpha + o(|t|^{\alpha+\gamma})$, as $|t| \to 0$ for some $0 < \alpha < 2, \gamma > 0$.
- R3: R is C^4 away from the origin and there exists a $c > 0$ such that $|R^{(4)}(t)| \leq ct^{\alpha-4}$

for all sufficiently small $t > 0$.

Define the class of C^1 diffeomorphisms to be the set of all continuous invertible maps $F : U \to \mathbb{R}^2$, such that F is $C^1(U)$ and F^{-1} is $C^1(V)$, where $V = F(U)$ is the range of F. By the inverse function theorem, necessary and sufficient conditions are that F be invertible, $C^1(U)$ and $\det(J_F) \neq 0$. We write C^1 as short for $C^1(\mathbb{R}^2)$. Moreover, the directional derivative in the direction θ, denoted $\partial_\theta f$, is $J_f^t u_\theta$, where $u_\theta = (\cos\theta, \sin\theta)$. So their basic assumption is that

$$F(\boldsymbol{t}+\boldsymbol{h}) = F(\boldsymbol{t}) + J_F^{\boldsymbol{t}}\boldsymbol{h} + o(|\boldsymbol{h}|)$$

where $J_F^{\boldsymbol{t}} := (\frac{\partial F_i}{\partial t_j}(t))_{i,j}$ is the Jacobian of the map F at $\boldsymbol{t}$.

They also note these consequences of a C^2 diffeomorphism assumption:

D1: F is a quasiconformal map on bounded simply connected domains.

D2: $\sup_{\boldsymbol{t}\in\Theta} |\frac{F(\boldsymbol{t}+\epsilon\boldsymbol{h})-F(\boldsymbol{t})}{\epsilon} - J_F^{\boldsymbol{t}}\boldsymbol{h}| \to 0$ as $\epsilon \to 0$ for every compact set Θ.

D3: For any vector $\boldsymbol{h} \neq 0$ and compact set Θ, there exists a constant c such that $|\partial_{\boldsymbol{h}}^{(2,2)} R(|F(\boldsymbol{s}) - F(\boldsymbol{t})|)| \leq c|\boldsymbol{s}-\boldsymbol{t}|^{\alpha-4}$ for all $\boldsymbol{s}, \boldsymbol{t} \in \Omega$ such that $\boldsymbol{s} \neq \boldsymbol{t}$.

D4: For every compact subset Θ, there exists constants $c_1, c_2 > 0$ such that $c_1|\boldsymbol{h}| \leq |J_F^{\boldsymbol{x}}\boldsymbol{h}| \leq c_2|\boldsymbol{h}|$ for all $\boldsymbol{h}$ and all $\boldsymbol{x} \in \Theta$.

D5: $|J_F^{\boldsymbol{t}}\boldsymbol{h}|^\alpha$ is Holder continuous in $\boldsymbol{t} \in \Theta$ for any h and compact set Θ.

Anderes and Chatterjee presented a class of deformation functions known as quasiconformal mappings and showed that can be estimated locally from estimates of $J_F^{\boldsymbol{t}}\boldsymbol{h}$ in directions $\boldsymbol{h} = (1, 0)$, (0,1) and (1,1).

D4 satisfies assumpton 2.3 (ii) and for assumpton 2.3 (i), Anderes and Chatterjee showed that if $\alpha < 2$

$$\begin{aligned} C(\boldsymbol{t}, \boldsymbol{t}+\boldsymbol{h}) &= R(|F(\boldsymbol{t}) - F(\boldsymbol{t}+\boldsymbol{h})|) \\ & R(0) - |J_F^{\boldsymbol{t}}\boldsymbol{h} + o(|\boldsymbol{h}|)|^\alpha + o(|J_F^{\boldsymbol{t}}\boldsymbol{h} + o(|\boldsymbol{h}|)|^{\alpha+\gamma}) \\ & R(0) - |J_F^{\boldsymbol{t}}\boldsymbol{h}|^\alpha + O(|\boldsymbol{h}|^{-\alpha-1}) + o(|\boldsymbol{h}|^{\alpha+\gamma}) \end{aligned}$$

implies that $\sigma^2 = R(0)$, $f_{\boldsymbol{t}}(\boldsymbol{h}) = |J_F^{\boldsymbol{t}}\frac{\boldsymbol{h}}{|\boldsymbol{h}|}|^\alpha$, $\alpha(\boldsymbol{t}) = \alpha$ and $\gamma(\boldsymbol{t}) = \min(\gamma, 1)$. Anderes and Chatterjee restrict α to be less than 2, however, if F is sufficiently smooth, we can consider estimation for $\alpha > 2$.

Here we will show that this model satisfies our covariance assumption.

Proof. Consider the stationary process X with isotropic covariance function

$$R(\boldsymbol{h}) = c_0 + \sum_{s=1}^{k+1} c_{2s}|\boldsymbol{h}|^{2s} + c_\alpha|\boldsymbol{h}|^\alpha + o(|\boldsymbol{h}|^{2k+3})$$

where $\alpha \in (2k, 2k+2)$. Suppose F is a deformation function from $\mathbb{R}^d$ to $\mathbb{R}^d$ that is $(k+1)$ times differentiable. Define $Y(\boldsymbol{t}) = X(F(\boldsymbol{t}))$. Write $F(\boldsymbol{t}) = (F_1(\boldsymbol{t}), \ldots, F_d(\boldsymbol{t}))'$ and recall the notation from section 1.2. Then

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

Thus, for $s = 1, \ldots, k+1$,

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

Replacing $\boldsymbol{h}$ with $\boldsymbol{t}-\boldsymbol{s}$, it is easily verified that this satisfies the first part of the covariance expansion in Assumption 2.3

Then let $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_d$ be some points on the line connecting $\boldsymbol{t}$ and $\boldsymbol{t}+\boldsymbol{h}$. By Taylor's theorem,

$$\begin{aligned}
&|F(\boldsymbol{t}+\boldsymbol{h}) - F(\boldsymbol{t})|^{\alpha} \\
= &\left(\sum_{j=1}^{d} \left(\sum_{|\boldsymbol{\ell}|=1}^{d} F_j^{(\boldsymbol{\ell})}(\boldsymbol{t})\boldsymbol{h}^{\boldsymbol{\ell}} + \frac{1}{2}\sum_{|\boldsymbol{\ell}|=2}^{d} F_j^{(\boldsymbol{\ell})}(\boldsymbol{z}_j)\boldsymbol{h}^{\boldsymbol{\ell}} \right)^2 \right)^{\alpha/2} \\
= &\left(\sum_{j=1}^{d} \left[\left(\sum_{|\boldsymbol{\ell}|=1}^{d} F_j^{(\boldsymbol{\ell})}(\boldsymbol{t})\boldsymbol{h}^{\boldsymbol{\ell}} \right)^2 + \frac{1}{2}\sum_{|\boldsymbol{\ell}|=1,2}^{d} F_j^{(\boldsymbol{\ell})}(\boldsymbol{t})\boldsymbol{h}^{\boldsymbol{\ell}} \sum_{|\boldsymbol{\ell}|=2}^{d} F_j^{(\boldsymbol{\ell})}(\boldsymbol{z}_j)\boldsymbol{h}^{\boldsymbol{\ell}} \right] \right)^{\alpha/2}.
\end{aligned}$$

Then by applying Taylor's theorem and letting $\lambda \in [0,1]$,

$$\begin{aligned}
&\left(\sum_{j=1}^{d} \left[\left(\sum_{|\boldsymbol{\ell}|=1}^{d} F_j^{(\boldsymbol{\ell})}(\boldsymbol{t})\boldsymbol{h}^{\boldsymbol{\ell}} \right)^2 + \frac{1}{2}\sum_{|\boldsymbol{\ell}|=1}^{d} F_j^{(\boldsymbol{\ell})}(\boldsymbol{t})\boldsymbol{h}^{\boldsymbol{\ell}} \sum_{|\boldsymbol{\ell}|=2}^{d} F_j^{(\boldsymbol{\ell})}(\boldsymbol{z}_j)\boldsymbol{h}^{\boldsymbol{\ell}} \right] \right)^{\alpha/2} \\
= &\left(\sum_{j=1}^{d} \left(\sum_{|\boldsymbol{\ell}|=1}^{d} F_j^{(\boldsymbol{\ell})}(\boldsymbol{t})\boldsymbol{h}^{\boldsymbol{\ell}} \right)^2 \right)^{\alpha/2} + \frac{\alpha}{4} \left(\sum_{j=1}^{d}\sum_{|\boldsymbol{\ell}|=1}^{d} F_j^{(\boldsymbol{\ell})}(\boldsymbol{t})\boldsymbol{h}^{\boldsymbol{\ell}} \sum_{|\boldsymbol{\ell}|=2}^{d} F_j^{(\boldsymbol{\ell})}(\boldsymbol{z}_j)\boldsymbol{h}^{\boldsymbol{\ell}} \right)
\end{aligned}$$

$$\cdot\left(\sum_{j=1}^{d}\left[\left(\sum_{|\boldsymbol{\ell}|=1}^{d}F_j^{(\boldsymbol{\ell})}(\boldsymbol{t})\boldsymbol{h}^{\boldsymbol{\ell}}\right)^2+\frac{\lambda}{2}\sum_{|\boldsymbol{\ell}|=1}^{d}F_j^{(\boldsymbol{\ell})}(\boldsymbol{t})\boldsymbol{h}^{\boldsymbol{\ell}}\sum_{|\boldsymbol{\ell}|=2}^{d}F_j^{(\boldsymbol{\ell})}(\boldsymbol{z}_j)\boldsymbol{h}^{\boldsymbol{\ell}}\right]\right)^{\alpha/2-1}$$

$$=\left(\sum_{j=1}^{d}\left(\sum_{|\boldsymbol{\ell}|=1}^{d}F_j^{(\boldsymbol{\ell})}(\boldsymbol{t})\boldsymbol{h}^{\boldsymbol{\ell}}\right)^2\right)^{\alpha/2}+O(|\boldsymbol{h}|^{1+\alpha}).$$

The last equality holds uniformly since F is twice continuously differentiable and applying the extreme value theorem. Therefore since $\alpha+1<2k+3$, we can write

$$C(\boldsymbol{t},\boldsymbol{t}+\boldsymbol{h})=c_0+\sum_{|\boldsymbol{\ell}|=2}^{2k+2}b_{\boldsymbol{\ell}}(\boldsymbol{t})\boldsymbol{h}^{\boldsymbol{\ell}}+|J_F^{\boldsymbol{t}}\boldsymbol{h}|^{\alpha}+O(|\boldsymbol{h}|^{2\nu+1})\text{ as }\boldsymbol{h}\to\boldsymbol{0}$$

for some functions $b_{\boldsymbol{\ell}}$. Using the fact that the covariance function is symmetric, it is straightforward to verify that this satisfies our covariance assumption with $\gamma(\boldsymbol{t})=1$ and $f_{\boldsymbol{t}}(\boldsymbol{h})=|J_f^{\boldsymbol{t}}\boldsymbol{h}|$. So therefore the deformation framework satisfies our covariance assumptions for general $\alpha(\boldsymbol{t})$ s.t. $\alpha(\boldsymbol{t})/2$ is non-integer if F is at least $\lfloor\alpha\rfloor$ times continuously differentiable.

The second half of the covariance assumption follows from the proof of Lemma 9 in Anderes and Chatterjee (2009). $\square$

Suppose now that the observed process Y is the deformation of a two dimensional isotropic Matérn Gaussian random field X with covariance $R(|\boldsymbol{t}|)$. Then

$$\begin{aligned}C(\boldsymbol{t},\boldsymbol{t}+\boldsymbol{h}) &= \mathrm{Cov}(Y(\boldsymbol{t}),Y(\boldsymbol{t}+\boldsymbol{h}))\\ &= \mathrm{Cov}\{X(F(\boldsymbol{t})),X(F(\boldsymbol{t}+\boldsymbol{h}))\}\\ &= R_\nu(|F(\boldsymbol{t}+\boldsymbol{h})-F(\boldsymbol{t})|)\\ &= \frac{\sigma^2\pi\nu^\nu}{\Gamma(\nu)\sin(\nu\pi)\rho^{2\nu}\Gamma(1+\nu)}\left\{\left|J_F^{\boldsymbol{t}}\frac{\boldsymbol{h}}{|\boldsymbol{h}|}\right|^{2\nu}+O(|\boldsymbol{h}|)\right\}.\end{aligned}$$

So the estimation procedure to recover F will actually estimate the function

$$G(\boldsymbol{t}):=\left(\frac{\sigma^2\pi\nu^\nu}{\Gamma(\nu)\sin(\nu\pi)\rho^{2\nu}\Gamma(1+\nu)}\right)^{1/(2\nu)}F(\boldsymbol{t})=BF(\boldsymbol{t}).$$

And so $G^{-1}(\boldsymbol{t})=F^{-1}\left(\frac{\boldsymbol{t}}{B}\right)$ and hence the covariance for $Y(G^{-1}(\boldsymbol{t}))$ is

$$C(G^{-1}(\boldsymbol{t}),G^{-1}(\boldsymbol{t}+\boldsymbol{h}))$$

$$
\begin{aligned}
&= \operatorname{Cov}\left\{X\left(F\left(F^{-1}\left(\frac{\boldsymbol{t}}{B}\right)\right)\right), X\left(F\left(F^{-1}\left(\frac{\boldsymbol{t}+\boldsymbol{h}}{B}\right)\right)\right)\right\} \\
&= \operatorname{Cov}\left\{X\left(\frac{\boldsymbol{t}}{B}\right), X\left(\frac{\boldsymbol{t}+\boldsymbol{h}}{B}\right)\right\} \\
&= R\left(\frac{\boldsymbol{h}}{B}\right).
\end{aligned}
$$

APPENDIX B

Proofs of Limit Theorems and Related Results

Preliminary Results

Lemma B.1. *Suppose Assumption 3.2 holds, $n \to \infty$ and $nb \to \infty$. Then*

$$\left| n^{-d} b^{-d-\sum_{\ell=1}^{d} m_\ell} \sum_{\boldsymbol{t}_i \in T_n(\boldsymbol{t}_0, b)} k_{\boldsymbol{t}_i} \prod_{\ell=1}^{d} t_{\boldsymbol{i}\ell}^{m_\ell} - \int_{[-\frac{\boldsymbol{t}_0}{b}, \frac{1-\boldsymbol{t}_0}{b}]} k(\boldsymbol{z}) \prod_{\ell=1}^{d} z_\ell^{m_\ell} d\boldsymbol{z} \right| \leq \mathbf{c}_{10} n^{-1} b^{-1},$$

for some $\mathbf{c}_{10} > 0$, i, j non-negative integers.

Proof. Let S_i denote the d-dim cube with area n^{-d} centered at $\boldsymbol{t}_i$. Thus, $\cup_i S_i = [0,1]^d$. We will also let $\sum_{\boldsymbol{t}}$ refer to $\sum_{\boldsymbol{t}_i \in T_n(\boldsymbol{t}_0, b)}$.

If we let $f(\boldsymbol{z}) = k(\boldsymbol{z}) \prod_{\ell=1}^{d} z_\ell^{m_\ell}$ then

$$\begin{aligned}
&\left| n^{-d} b^{-d-\sum_{\ell=1}^{d} m_\ell} \sum_{\boldsymbol{t}_i} k_{\boldsymbol{t}_i} \prod_{\ell=1}^{d} t_{\boldsymbol{i}\ell}^{m_\ell} - \int_{[-\frac{\boldsymbol{t}_0}{b}, \frac{1-\boldsymbol{t}_0}{b}]} f(\boldsymbol{z}) d\boldsymbol{z} \right| \\
\leq \;& b^{-d} \left| n^{-d} \sum_{\boldsymbol{t}_i} f\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right) - \sum_{\boldsymbol{t}_i} \int_{S_i} f\left(\frac{\boldsymbol{s} - \boldsymbol{t}_0}{b}\right) d\boldsymbol{s} \right| && \text{(B.1)} \\
+ & \left| b^{-d} \sum_{\boldsymbol{t}_i} \int_{S_i} f\left(\frac{\boldsymbol{s} - \boldsymbol{t}_0}{b}\right) d\boldsymbol{s} - \int_{[-\frac{\boldsymbol{t}_0}{b}, \frac{1-\boldsymbol{t}_0}{b}]} f(\boldsymbol{z}) d\boldsymbol{z} \right|. && \text{(B.2)}
\end{aligned}$$

For (B.1),

$$\left| f\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right) - f\left(\frac{\boldsymbol{s} - \boldsymbol{t}_0}{b}\right) \right| \le \left| \left(\frac{\boldsymbol{s} - \boldsymbol{t}_i}{b}\right)' f'\left(\frac{\boldsymbol{z} - \boldsymbol{t}_0}{b}\right) \right|$$

for some $\boldsymbol{z}$ between $\boldsymbol{s}$ and $\boldsymbol{t}_i$ and $f'(\boldsymbol{z}) = \nabla f(\boldsymbol{z})$. Let $\|f'\|_\infty = \max_{\boldsymbol{z}\in[-1,1]^d} |f'(\boldsymbol{z})|$, which is bounded by Assumption 3.2. Then since $\left|\frac{\boldsymbol{s}-\boldsymbol{t}_i}{b}\right| \le n^{-1}b^{-1}$,

$$\begin{aligned} & b^{-d}\left| n^{-d}\sum_{\boldsymbol{t}_i} f\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right) - \sum_{\boldsymbol{t}_i}\int_{S_i} f\left(\frac{\boldsymbol{s} - \boldsymbol{t}_0}{b}\right) d\boldsymbol{s} \right| \\ \le \;& n^{-d}b^{-d}\sum_{\boldsymbol{t}_i} n^{-1}b^{-1}\|f'\|_\infty \le 2^d n^{-1}b^{-1}\|f'\|_\infty. \end{aligned}$$

For (B.2), letting $\boldsymbol{z} = (\boldsymbol{s} - \boldsymbol{t}_0)/b$,

$$\begin{aligned} b^{-d}\sum_{\boldsymbol{t}_i}\int_{S_i} f\left(\frac{\boldsymbol{s} - \boldsymbol{t}_0}{b}\right) d\boldsymbol{s} &= b^{-d}\int_{[0,1]^d} f\left(\frac{\boldsymbol{s} - \boldsymbol{t}_0}{b}\right) d\boldsymbol{s} \\ &= \int_{[-\frac{\boldsymbol{t}_0}{b}, \frac{1-\boldsymbol{t}_0}{b}]} f(\boldsymbol{z})d\boldsymbol{z}. \end{aligned}$$

Since the support of f is $[0,1]^d$, this integral is effectively

$$\int_{[-(1\wedge\frac{\boldsymbol{t}_0}{b}), \frac{1-\boldsymbol{t}_0}{b}\wedge 1]} f(\boldsymbol{z})d\boldsymbol{z},$$

but the former notation is simpler. So

$$\left| n^{-d}b^{-d-\sum_{\ell=1}^d m_\ell}\sum_{\boldsymbol{t}_i} k_{\boldsymbol{t}_i}\prod_{\ell=1}^d (t_{i\ell} - t_{0\ell})^{m_\ell} - \int_{[-\frac{\boldsymbol{t}_0}{b}, \frac{1-\boldsymbol{t}_0}{b}]} f(\boldsymbol{z})d\boldsymbol{z} \right| \le \mathbf{c}_{10}n^{-1}b^{-1}$$

for some $\mathbf{c}_{10} > 0$. □

Theorem B.2 (Gaussian Moment Theorem)**.** *Suppose $X_1, X_2, ..., X_n$ are jointly Gaussian with mean 0 and finite variance. Then for n odd,*

$$E\left(\prod_{i=1}^n X_i\right) = 0$$

and for n even,

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

where the sum is over the $\frac{n}{2}$ unique pairings that can be chosen from $1, 2, ..., n$.

Theorem B.3. *Suppose m, n are non-negative integers and at least one is non-zero, and $X_1, X_2, ..., X_m, X_{m+1}, ..., X_{m+n}$ are jointly Gaussian with mean 0. Then for $m = 1, n = 0$ or $n > 0$ odd, we have*

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

and for $n > 0$ even,

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

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

Proof. The proof is by induction on m. Clearly the result is true for $m = 0$ and any $n > 0$ or when $m = 1$ and $n = 0$. Suppose the result is true for finite m and any n. Then for $m + 1$,

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

and by the induction hypothesis, the above is 0 if n is odd, and it is equal to the desired expression if n is even. Therefore we have shown that the result is true for any m and n where at least one is non-zero. □

Corollary B.4. *Suppose $X_1, X_2, ..., X_n$ are jointly Gaussian. Then*

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

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

Lemma B.5. *Let Assumptions 3.1 and 3.2 hold. Then if* $\boldsymbol{t} \in T_n(\boldsymbol{t}_0, b)$,

$$\sum_{\boldsymbol{t}_i \in T_n(\boldsymbol{t}_0, b)} |C_n(\boldsymbol{t}, \boldsymbol{t}_i)| \leq \mathbf{c}_{11} r_{nb}(\boldsymbol{t}), \tag{B.3}$$

uniformly over $\boldsymbol{t}_0 \in [0,1]^d$ *for some* $\mathbf{c}_{11} > 0$ *where*

$$r_{nb}(\boldsymbol{t}) = \begin{cases} 1 & \text{if } \psi(\boldsymbol{t}) > d \\ \log(nb) & \text{if } \psi(\boldsymbol{t}) = d \\ (nb)^{d-\psi(\boldsymbol{t}_0)} & \text{if } \psi(\boldsymbol{t}) < d. \end{cases}$$

Proof. From Assumption 3.1, $|C_n(\boldsymbol{t}, \boldsymbol{t} + \boldsymbol{j}/n)| \leq \mathbf{c}_5 \; \forall \boldsymbol{j} \in \mathbb{Z}^d$ and $|C_n(\boldsymbol{t}, \boldsymbol{t} + \boldsymbol{j}/n)| \leq \mathbf{c}_6 |\boldsymbol{j}|^{\psi(\boldsymbol{t})}$ for $|\boldsymbol{j}| > \tau$. So

$$\begin{aligned}
\sum_{\boldsymbol{t}_i \in T_n(\boldsymbol{t}_0, b)} |C_n(\boldsymbol{t}, \boldsymbol{t}_i)| &= \sum_{|\boldsymbol{t}_i - \boldsymbol{t}| \leq \tau/n} |C_n(\boldsymbol{t}, \boldsymbol{t}_i)| + \sum_{|\boldsymbol{t}_i - \boldsymbol{t}| > \tau/n} |C_n(\boldsymbol{t}, \boldsymbol{t}_i)| \\
&\leq \sum_{\boldsymbol{j} \in \mathbb{Z}^d \cap \Omega_{0,\tau}} |C_n(\boldsymbol{t}, \boldsymbol{t} + \boldsymbol{j}/n)| + \sum_{\boldsymbol{j} \in (\mathbb{Z}^2 \cap \Omega_{0,\tau})^c, |j_i| < 2nb} |C_n(\boldsymbol{t}, \boldsymbol{t} + \boldsymbol{j}/n)| \\
&\leq (2\tau)^d \mathbf{c}_5 + \sum_{\boldsymbol{j} \in (\mathbb{Z}^2 \cap \Omega_{0,\tau})^c, |j_i| 3nb} \mathbf{c}_6 |\boldsymbol{j}|^{-\psi(\boldsymbol{t})} \\
&\leq (2\tau)^d \mathbf{c}_5 + \sum_{j=1}^{3nb} 3^d j^{d-1-\psi(\boldsymbol{t})} \\
&\leq (2\tau)^d \mathbf{c}_5 + \int_{j=1}^{3nb} 3^d j^{d-1-\psi(\boldsymbol{t})} \\
&= (2\tau)^d \mathbf{c}_5 + (d - \psi(\boldsymbol{t})) 3^d [(3nb)^{d-\psi(\boldsymbol{t})} - 1].
\end{aligned}$$

So therefore the lemma is satisfied for

$$\mathbf{c}_{11} = \begin{cases} (2\tau)^d \mathbf{c}_5 + (\psi(\boldsymbol{t}) - d) 3^d, & \text{if } \psi(\boldsymbol{t}) > d \\ (2\tau)^d \mathbf{c}_5 + (d - \psi(\boldsymbol{t})) 3^d + \log(3), & \text{if } \psi(\boldsymbol{t}) = d \\ (2\tau)^d \mathbf{c}_5 + (d - \psi(\boldsymbol{t})) 3^d 3^{d-\psi(\boldsymbol{t})}, & \text{if } \psi(\boldsymbol{t}) < d \end{cases}.$$

□

Theorem B.6. *Suppose* $X, X_1, X_2, ...$ *are random variables where the moments of* X *all exist and are finite and uniquely determine its distribution, and* $E(X_n^k) \to E(X^k)$

for all $k = 1, 2, \ldots$. Then

$$X_n \xrightarrow{D} X$$

Proof. Now $\limsup_{n\to\infty} E(X_n^2) = E(X^2) < \infty$ implies that $\{X_n\}$ is tight. So suppose that X_n' is a subsequence converging in distribution to Y. By Prokhorov, to complete the proof it suffices to show that Y has the same distribution as X'. By a standard result, $X_{n'}^k \xrightarrow{D} X^k$, and for any positive integer k, $X_{n'}^k$ is uniformly integrable since $\limsup_{n'\to\infty} E(X_{n'}^{k+1}) = E(X^{k+1}) < \infty$. Thus

$$E(X_{n'}^k) \to E(Y^k)$$

and so Y has the same moments as X, and thus must have the same distribution. □

Notationally, let $\widetilde{W}_n(\boldsymbol{t}) = W_n(\boldsymbol{t}) - \mu_n(\boldsymbol{t})$ and let $\sum_{\boldsymbol{t}_i}$ refer to $\sum_{\boldsymbol{t}_i \in T_n(\boldsymbol{t}_0, b)}$.

Theorem B.7. *Let Assumptions 3.1(iii),(iv) and 3.6 hold. Then for some $\mathbf{c}_{12} > 0$,*

$$\frac{(nb)^{d/2}}{\log(n)} \sup_{\boldsymbol{t}_0 \in [0,1]^2} \left| \frac{1}{n^d b^d} \sum_{\boldsymbol{t}_i} k\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right) (\widetilde{W}_n(\boldsymbol{t}_i)^2 - E\widetilde{W}_n(\boldsymbol{t}_i)^2) \right| \le \mathbf{c}_{12}$$

eventually with probability 1 as $n \to \infty$ and $nb \to \infty$.

Proof. Let $\hat{\beta}_0(\boldsymbol{t}_0) = \frac{1}{n^d b^d} \sum_{\boldsymbol{t}_i} k\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right) \widetilde{W}_n^2(\boldsymbol{t}_i)$ where k satisfies Assumption 3.6. From Assumption 3.6, $\int_{-1}^1 |k_i'(t)| dt < \infty$ for $i = 1, 2, \ldots, d$ and so k_i has bounded variation. Then we can write k_i as $k_{1,i} - k_{2,i}$ where $k_{1,i}$ and $k_{2,i}$ are monotone increasing. Define

$$\hat{\beta}_{\boldsymbol{j}}(\boldsymbol{t}_0) = \frac{1}{n^d b^d} \sum_{\boldsymbol{t}_i} \prod_{\ell=1}^{d} k_{j_\ell,\ell}(t_{i\ell} - t_{0\ell}) \widetilde{W}_n^2(\boldsymbol{t}_i),$$

for $j_\ell \in \{1, 2\}$, $\ell = 1, \cdots, d$. Let $\mathbf{1}(\cdot)$ denote the indicator function. Then

$$\begin{aligned}
& \frac{1}{n^d b^d} \sum_{\boldsymbol{t}_i} \prod_{\ell=1}^{d} k_{j_\ell,\ell}(t_{i\ell} - t_{0\ell}) \widetilde{W}_n^2(\boldsymbol{t}_i) \\
= \; & \frac{1}{n^d b^d} \sum_{\boldsymbol{t}_i} \mathbf{1}(-b \le \boldsymbol{t}_i - \boldsymbol{t}_0 \le b) \prod_{\ell=1}^{d} \int_{-b}^{t_{i\ell} - t_{0\ell}} dk_{j_\ell,\ell}(v_\ell) \widetilde{W}_n^2(\boldsymbol{t}_i) \\
= \; & \frac{1}{n^d b^d} \int_{[-b,b]^d} \sum_{\boldsymbol{t}_i} \widetilde{W}_n^2(\boldsymbol{t}_i) \mathbf{1}(\boldsymbol{v} \le \boldsymbol{t}_i - \boldsymbol{t}_0 \le b) dk_{j_1,1}(v_1) \cdots dk_{j_d,d}(v_d)
\end{aligned}$$

$$= \frac{1}{n^d b^d} \int\limits_{[-b,b]^d} \sum_{\boldsymbol{t}_i} \widetilde{W}_n^2(\boldsymbol{t}_i) \mathbf{1}(\boldsymbol{t}_0 + \boldsymbol{v} \leq \boldsymbol{t}_i \leq \boldsymbol{v} + b) dk_{j_1,1}(v_1) \cdots dk_{j_d,d}(v_d),$$

where $\boldsymbol{t} \leq b$ is evaluated component-wise, i.e. $t_\ell \leq b$ for $i = 1, \cdots, \ell$. The interchange of summation and integration is justified by the fact that this is a Stieltjes integral since $k_{j_\ell,\ell}$ are monotone increasing. Now let

$$\begin{aligned} G_n(\boldsymbol{u}, \boldsymbol{v}) &= \frac{1}{n^d b^d} \sum_{\boldsymbol{t}_i \in T_n} \widetilde{W}_n^2(\boldsymbol{t}_i) \mathbf{1}(\boldsymbol{u} \leq \boldsymbol{t}_i \leq \boldsymbol{v}) \mathbf{1}(0 \leq \boldsymbol{t}_i \leq 1), \\ G(\boldsymbol{u}, \boldsymbol{v}) &= E G_n(\boldsymbol{u}, \boldsymbol{v}). \end{aligned}$$

Then

$$\begin{aligned} & \hat{\beta}_{\boldsymbol{j}}(\boldsymbol{t}_0) - E\hat{\beta}_{\boldsymbol{j}}(\boldsymbol{t}_0) \\ = & \int\limits_{[-b,b]^d} \left(G_n(\boldsymbol{t}_0 + \boldsymbol{v}, \boldsymbol{t}_0 + b) - G(\boldsymbol{t}_0 + \boldsymbol{v}, \boldsymbol{t}_0 + b)\right) dk_{j_1,1}(v_1) \cdots dk_{j_d,d}(v_d). \end{aligned}$$

So now,

$$\sup_{\boldsymbol{t}_0 \in [0,1]^d} |\hat{\beta}_{\boldsymbol{j}}(\boldsymbol{t}) - E\hat{\beta}_{\boldsymbol{j}}(\boldsymbol{t}_0)| \leq \sup_{\boldsymbol{t}_0 \in [0,1]^d} \sup_{0 \leq \boldsymbol{u} \leq 2b} |G_n(\boldsymbol{t}_0, \boldsymbol{t}_0 + \boldsymbol{u}) - G(\boldsymbol{t}_0, \boldsymbol{t}_0 + \boldsymbol{u})| \prod_{\ell=1}^{d} \int_{-b}^{u_\ell} dk_{j_\ell,\ell}.$$

We will prove the convergence of

$$\sup_{\boldsymbol{t}_0 \in [0,1]^d} \sup_{0 \leq \boldsymbol{u} \leq 2b} |G_n(\boldsymbol{t}_0, \boldsymbol{t}_0 + \boldsymbol{u}) - G(\boldsymbol{t}_0, \boldsymbol{t}_0 + \boldsymbol{u})|$$

by discretizing $[0,1]^d$. Let $\mathcal{G}$ be a grid of size $2b$ on $[0,1]^d$ where each grid point is the lower-left corner (an "anchor point") of a grid cell with width less than or equal to $2b$. Then for $\boldsymbol{t}_0 \in [0,1]^d$, the rectangle defined by the points $\boldsymbol{t}_0$ and $\boldsymbol{t}_0 + \boldsymbol{u}$ can intersect with up to 2^d grid cells with anchor points $\tilde{\boldsymbol{t}}_1, ..., \tilde{\boldsymbol{t}}_j$, $j \leq 2^d$. Therefore,

$$|G_n(\boldsymbol{t}_0, \boldsymbol{t}_0 + \boldsymbol{u}) - G(\boldsymbol{t}_0, \boldsymbol{t}_0 + \boldsymbol{u})| \leq 2^{2d} \max_i \sup_{0 \leq \boldsymbol{u} \leq 2b} |G_n(\tilde{\boldsymbol{t}}_i, \tilde{\boldsymbol{t}}_i + \boldsymbol{u}) - G(\tilde{\boldsymbol{t}}_i, \tilde{\boldsymbol{t}}_i + \boldsymbol{u})|$$

and further

$$\begin{aligned} & \sup_{\boldsymbol{t}_0 \in [0,1]^d} \sup_{0 \leq \boldsymbol{u} \leq 2b} |G_n(\boldsymbol{t}_0, \boldsymbol{t}_0 + \boldsymbol{u}) - G(\boldsymbol{t}_0, \boldsymbol{t}_0 + \boldsymbol{u})| \\ \leq & \ 2^{2d} \sup_{\tilde{\boldsymbol{t}} \in \mathcal{G}} \sup_{0 \leq \boldsymbol{u} \leq 2b} |G_n(\tilde{\boldsymbol{t}}, \tilde{\boldsymbol{t}} + \boldsymbol{u}) - G(\tilde{\boldsymbol{t}}, \tilde{\boldsymbol{t}} + \boldsymbol{u})|. \qquad \text{(B.4)} \end{aligned}$$

Assume that the point $(0,0)$ is always a grid point in $\mathcal{G}$. Next, fix $\boldsymbol{t}$ and $\boldsymbol{u}$ and suppose that $\boldsymbol{t}+\boldsymbol{u}$ falls in the grid cell of $\mathcal{G}$ with anchor point $\tilde{\boldsymbol{t}}$. Then we discretize this grid cell by considering the points on T_n which also fall within this cell and denoted as $\mathcal{H}$. Assume that $\boldsymbol{t}+\boldsymbol{u}$ falls within the grid cell of $\mathcal{H}$ with lower-left corner $\boldsymbol{t}+\tilde{\boldsymbol{u}}$. Then

$$G_n(\boldsymbol{t},\boldsymbol{t}+\boldsymbol{u}) - G(\boldsymbol{t},\boldsymbol{t}+\boldsymbol{u}) \;\; = \;\; G_n(\boldsymbol{t},\boldsymbol{t}+\tilde{\boldsymbol{u}}) - G(\boldsymbol{t},\boldsymbol{t}+\tilde{\boldsymbol{u}}).$$

So therefore

$$\sup_{\tilde{\boldsymbol{t}}\in\mathcal{G}}\sup_{0\le\boldsymbol{u}\le 2b}|G_n(\tilde{\boldsymbol{t}},\tilde{\boldsymbol{t}}+\boldsymbol{u}) - G(\tilde{\boldsymbol{t}},\tilde{\boldsymbol{t}}+\boldsymbol{u})| = \sup_{\tilde{\boldsymbol{t}}\in\mathcal{G}}\sup_{\boldsymbol{v}\in\mathcal{H}}|G_n(\tilde{\boldsymbol{t}},\tilde{\boldsymbol{t}}+\boldsymbol{v}) - G(\tilde{\boldsymbol{t}},\tilde{\boldsymbol{t}}+\boldsymbol{v})| \quad \text{(B.5)}$$

By combining (B.4) and (B.5),

$$\begin{aligned}
&\frac{(nb)^{d/2}}{\log(n)}\sup_{\boldsymbol{t}_0\in[0,1]^2}\sup_{\boldsymbol{u}\le 2b,\boldsymbol{u}>0}|G_n(\boldsymbol{t}_0,\boldsymbol{t}_0+\boldsymbol{u}) - G(\boldsymbol{t}_0,\boldsymbol{t}_0+\boldsymbol{u})| \\
\le\;& \frac{2^{2d}(nb)^{d/2}}{\log(n)}\sup_{\tilde{\boldsymbol{t}}\in\mathcal{G}}\sup_{\boldsymbol{v}\in\mathcal{H}}|G_n(\tilde{\boldsymbol{t}},\tilde{\boldsymbol{t}}+\boldsymbol{v}) - G(\tilde{\boldsymbol{t}},\tilde{\boldsymbol{t}}+\boldsymbol{v})|. \qquad \text{(B.6)}
\end{aligned}$$

Then for (B.6), from Corollary 3.7 and the Markov inequality,

$$\begin{aligned}
&P\left(\frac{2^{2d}(nb)^{d/2}}{\log(n)}|G_n(\boldsymbol{t}_i,\boldsymbol{t}_i+\boldsymbol{v}_j) - G(\boldsymbol{t}_i,\boldsymbol{t}_i+\boldsymbol{v}_j)| > (d+2)\mathbf{c}_{13}\right) \\
=\;& P\left(\exp\left\{\frac{2^{2d}(nb)^{d/2}}{\mathbf{c}_{13}}|G_n(\boldsymbol{t}_i,\boldsymbol{t}_i+\boldsymbol{v}_j) - G(\boldsymbol{t}_i,\boldsymbol{t}_i+\boldsymbol{v}_j)|\right\} > e^{(d+2)\log(n)}\right) \\
\le\;& e^{-(d+2)\log(n)}\sum_{x=0}^{\infty}\frac{1}{x!}E\left(\frac{2^{2d}(nb)^{d/2}}{\mathbf{c}_{13}}\{|G_n(\boldsymbol{t}_i,\boldsymbol{t}_i+\boldsymbol{v}_j) - G(\boldsymbol{t}_i,\boldsymbol{t}_i+\boldsymbol{v}_j)|\}\right)^x \\
\le\;& e^{-(d+2)\log(n)}\sum_{x=0}^{\infty}\frac{1}{x!}\left(\frac{2^{2d}}{\mathbf{c}_{13}}\right)^x(2x-1)!!\mathbf{c}_7^x \\
\le\;& e^{-(d+2)\log(n)} \\
=\;& n^{-(d+2)}
\end{aligned}$$

for sufficiently large $\mathbf{c}_{13} > 0$. So now,

$$\begin{aligned}
&\sum_{n=1}^{\infty}P\left(\sup_{\boldsymbol{t}\in\mathcal{G}}\sup_{\boldsymbol{v}\in\mathcal{H}}\frac{2^{2d}(nb)^{d/2}}{\log(n)}|G_n(\boldsymbol{t},\boldsymbol{t}+\boldsymbol{v}) - G(\boldsymbol{t},\boldsymbol{t}+\boldsymbol{v})| > (d+2)\mathbf{c}_{13}\right) \\
\le\;& \sum_{n=1}^{\infty}\sum_{\boldsymbol{t}_i\in\mathcal{G}}\sum_{\boldsymbol{v}_j\in\mathcal{H}}P\left(\frac{2^{2d}(nb)^{d/2}}{\log(n)}|G_n(\boldsymbol{t}_i,\boldsymbol{t}_i+\boldsymbol{v}_j) - G(\boldsymbol{t}_i,\boldsymbol{t}_i+\boldsymbol{v}_j)| > (d+2)\mathbf{c}_{13}\right)
\end{aligned}$$

$$\begin{aligned}
&\leq \sum_{n=1}^{\infty} \sum_{\boldsymbol{t}_i \in \mathcal{G}} \sum_{\boldsymbol{v}_j \in \mathcal{H}} n^{-(d+2)} \\
&= \sum_{n=1}^{\infty} n^{-2} < \infty.
\end{aligned}$$

Applying the Borel-Cantelli lemma we have that for some $\mathbf{c}_{14} > 0$,

$$\frac{(nb)^{d/2}}{\log(n)} \sup_{\boldsymbol{t} \in [0,1]^d} \sup_{\boldsymbol{u} \leq 2b, \boldsymbol{u} > 0} |G_n(\boldsymbol{t}, \boldsymbol{t}+\boldsymbol{u}) - G(\boldsymbol{t}, \boldsymbol{t}+\boldsymbol{u})| \leq \mathbf{c}_{14}$$

eventually w.p.1. Therefore, with $\|k_\ell\|_\infty < \infty$, $\ell \leq d$,

$$\begin{aligned}
& \frac{(nb)^{d/2}}{\log(n)} \sup_{\boldsymbol{t} \in [0,1]^d} |\hat{\beta}(\boldsymbol{t}) - E\hat{\beta}(\boldsymbol{t})| \\
\leq\ & \frac{(nb)^{d/2}}{\log(n)} \sup_{\boldsymbol{t} \in [0,1]^d} \sum_{\boldsymbol{j} \in \{0,1\}^d} |\hat{\beta}_{\boldsymbol{j}}(\boldsymbol{t}) - E\hat{\beta}_{\boldsymbol{j}}(\boldsymbol{t})| \\
\leq\ & \frac{(nb)^{d/2}}{\log(n)} \sup_{\boldsymbol{t} \in [0,1]^d} \sup_{0 \leq \boldsymbol{u} \leq 2b} \sum_{\boldsymbol{j} \in \{0,1\}^d} |G_n(\boldsymbol{t}, \boldsymbol{t}+\boldsymbol{u}) - G(\boldsymbol{t}, \boldsymbol{t}+\boldsymbol{u})| \prod_{\ell=1}^{d} \int_{-b}^{u_\ell} dk_{j_\ell,\ell} \\
\leq\ & \frac{2^d (nb)^{d/2}}{\log(n)} \prod_{\ell=1}^{d} \|k_\ell\|_\infty \sup_{\boldsymbol{t} \in [0,1]^d} \sup_{0 \leq \boldsymbol{u} \leq 2b} |G_n(\boldsymbol{t}, \boldsymbol{t}+\boldsymbol{u}) - G(\boldsymbol{t}, \boldsymbol{t}+\boldsymbol{u})| \\
\leq\ & 2^d \prod_{\ell=1}^{d} \|k_\ell\|_\infty \mathbf{c}_{14}
\end{aligned}$$

eventually w.p.1. $\square$

Corollary B.8. *Let Assumptions 3.1(iii), (iv) and 3.6 hold. Then for some* $\mathbf{c}_{12} > 0$,

$$\frac{(nb)^{d/2}}{\log(n)} \sup_{\boldsymbol{t}_0 \in [0,1]^2} \left| \frac{1}{n^d b^d r_{nb}(\boldsymbol{t}_0)^{1/2}} \sum_{\boldsymbol{t}_i} k\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right) \widetilde{W}_n(\boldsymbol{t}_i) \right| \leq \mathbf{c}_{12}$$

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

Proof. Recall that $\widetilde{W}_n(\boldsymbol{t}_i) \sim N(0, C_n(\boldsymbol{t}_i, \boldsymbol{t}_i))$. This implies that

$$\frac{1}{(nb)^d} \sum_{\boldsymbol{t}_i} k_{\boldsymbol{t}_i} \widetilde{W}_n(\boldsymbol{t}_i) \sim N\left(0, \frac{1}{(nb)^{2d}} \sum_{\boldsymbol{t}_i, \boldsymbol{t}_j} k_{\boldsymbol{t}_i} k_{\boldsymbol{t}_j} E[\widetilde{W}_n(\boldsymbol{t}_i)\widetilde{W}_n(\boldsymbol{t}_j)]\right)$$

and further from lemma B.5,

$$\left| \frac{1}{(nb)^{2d}} \sum_{\boldsymbol{t}_i, \boldsymbol{t}_j} k_{\boldsymbol{t}_i} k_{\boldsymbol{t}_j} E[\widetilde{W}_n(\boldsymbol{t}_i) \widetilde{W}_n(\boldsymbol{t}_j)] \right| \leq \frac{\mathbf{c}_{11} r(nb)}{(nb)^d}.$$

Therefore since $\widetilde{W}_n$ is Gaussian,

$$E \left| \frac{1}{(nb)^d} \sum_{\boldsymbol{t}_i} k_{\boldsymbol{t}_i} \widetilde{W}_n(\boldsymbol{t}_i) \right|^x \leq \frac{x!! \mathbf{c}_{11}^{x/2} r(nb)^{x/2}}{(nb)^{dx/2}}$$

and the corollary follows from the proof of Theorem B.7 □

Proofs of the Main Results in Chapters 2 and 3

Proof of lemma 2.8

Proof. First we will prove (i).

$$\begin{aligned} & E\Delta^x_{\boldsymbol{h}/n}(Y(\boldsymbol{t}) - \mu(\boldsymbol{t}))\Delta^x_{\boldsymbol{h}/n}(Y(\boldsymbol{t} + \boldsymbol{u}/n) - \mu(\boldsymbol{t} + \boldsymbol{u}/n)) \\ = & \sum_{i=0}^{x} \sum_{j=0}^{x} (-1)^{i+j} \binom{x}{i} \binom{x}{j} C(\boldsymbol{t} + i\boldsymbol{h}/n, \boldsymbol{t} + \boldsymbol{u}/n + j\boldsymbol{h}/n), \end{aligned}$$

so we if we let $\tilde{\alpha}(\boldsymbol{t}) = (\alpha(\boldsymbol{t}) + \alpha(\boldsymbol{t} + \boldsymbol{u}/n)/2$, we can write

$$\begin{aligned} & \left| n^{\tilde{\alpha}(\boldsymbol{t})} E\Delta^x_{\boldsymbol{h}/n} \big(Y(\boldsymbol{t}) - \mu(\boldsymbol{t}))\Delta^x_{\boldsymbol{h}/n}(Y(\boldsymbol{t} + \boldsymbol{u}/n) - \mu(\boldsymbol{t} + \boldsymbol{u}/n)\big) - J_{\boldsymbol{t}}(\mathbf{u}) \right| \\ \leq \quad & |e_1| + |e_2| + |e_3| + |e_4| + |e_5|, \end{aligned}$$

where

$$\begin{aligned} e_1 &= n^{\tilde{\alpha}(\boldsymbol{t})} \sum_{i=0}^{x} \sum_{j=0}^{x} (-1)^{i+j} \binom{x}{i} \binom{x}{j} \sum_{|\boldsymbol{\ell}|=0}^{r} b_{\boldsymbol{\ell}}(\boldsymbol{t} + i\boldsymbol{h}/n)(\boldsymbol{t} + \boldsymbol{u}/n + j\boldsymbol{h}/n)^{\boldsymbol{\ell}} \\ e_2 &= n^{\tilde{\alpha}(\boldsymbol{t})} \sum_{i=0}^{x} \sum_{j=0}^{x} (-1)^{i+j} \binom{x}{i} \binom{x}{j} \sum_{|\boldsymbol{\ell}|=0}^{r} b_{\boldsymbol{\ell}}(\boldsymbol{t} + \boldsymbol{u} + j\boldsymbol{h}/n)(\boldsymbol{t} + i\boldsymbol{h}/n)^{\boldsymbol{\ell}} \\ e_3 &= n^{\tilde{\alpha}(\boldsymbol{t})} \sum_{i=0}^{x} \sum_{j=0}^{x} (-1)^{i+j} \binom{x}{i} \binom{x}{j} \\ & \quad \cdot \left\{ f_{\boldsymbol{t}+i\boldsymbol{h}/n}(\boldsymbol{u} + (i-j)\boldsymbol{h}) - f_{\boldsymbol{t}}(\boldsymbol{u} + (i-j)\boldsymbol{h}) \right\} |\boldsymbol{u}/n + (i-j)\boldsymbol{h}/n|^{\alpha(\boldsymbol{t})} \end{aligned}$$

$$
\begin{aligned}
e_4 &= n^{\tilde{\alpha}(\boldsymbol{t})} \sum_{i=0}^{x} \sum_{j=0}^{x} (-1)^{i+j} \binom{x}{i} \binom{x}{j} f_{\boldsymbol{t}+i\boldsymbol{h}/n}(\boldsymbol{u} + (i-j)\boldsymbol{h}) \\
&\quad \cdot \left\{ |\boldsymbol{u}/n + (i-j)\boldsymbol{h}/n|^{\alpha(\boldsymbol{t}+i\boldsymbol{h}/n)} - |\boldsymbol{u}/n + (i-j)\boldsymbol{h}/n|^{\alpha(\boldsymbol{t})} \right\} \\
e_5 &= n^{\tilde{\alpha}(\boldsymbol{t})} \sum_{i=0}^{x} \sum_{j=0}^{x} (-1)^{i+j} \binom{x}{i} \binom{x}{j} O\left(\boldsymbol{u}/n + (i-j)|\boldsymbol{h}/n|^{\alpha(\boldsymbol{t}+i\boldsymbol{h}/n)+\gamma(\boldsymbol{t}+i\boldsymbol{h}/n)} \right).
\end{aligned}
$$

First notice that

$$
\sum_{i=0}^{x} (-1)^i \binom{x}{i} i^k = 0 \tag{B.7}
$$

for any integer $k < x$. So it immediately follows that e_1 and e_2 are 0.

Let $\nabla f_{\boldsymbol{t}}(\boldsymbol{h})$ denote the gradient of f w.r.t. $\boldsymbol{t}$. Recall that $f_{\boldsymbol{t}}(\boldsymbol{h})$ is differentiable $\boldsymbol{t}$, so by Taylor's theorem,

$$
f_{\boldsymbol{t}+\boldsymbol{j}/n}(\boldsymbol{h}) - f_{\boldsymbol{t}}(\boldsymbol{h}) = n^{-1}\boldsymbol{j} \cdot \nabla f_{\boldsymbol{z}}(\boldsymbol{h}) \tag{B.8}
$$

for some $\boldsymbol{z}$ on the line connecting $\boldsymbol{t}$ and $\boldsymbol{t}+\boldsymbol{j}/n$. Then by the extreme value theorem, since $[0,1]^2$ is closed and f has continuous first order partial derivatives w.r.t $\boldsymbol{t}$,

$$
\sup_{\boldsymbol{z}\in[0,1]^d, |\boldsymbol{j}|=1, |\boldsymbol{h}|=1} |\boldsymbol{j} \cdot \nabla f_{\boldsymbol{z}}(\boldsymbol{h})| \leq \mathbf{c}_{15}
$$

for some $\mathbf{c}_{15} > 0$. Since $\alpha(\boldsymbol{t})$ is continuously differentiable by assumption,

$$
\begin{aligned}
\left|\frac{\boldsymbol{h}}{n}\right|^{\alpha(\boldsymbol{t})} - \left|\frac{\boldsymbol{h}}{n}\right|^{\alpha(\boldsymbol{t}+\boldsymbol{j}/n)} &= \left|\frac{\boldsymbol{h}}{n}\right|^{\alpha(\boldsymbol{t})} \left(1 - \left|\frac{\boldsymbol{h}}{n}\right|^{\alpha(\boldsymbol{t}+\boldsymbol{j}/n)-\alpha(\boldsymbol{t})} \right) \\
&= \left|\frac{\boldsymbol{h}}{n}\right|^{\alpha(\boldsymbol{t})} \left(1 - \left|\frac{\boldsymbol{h}}{n}\right|^{\boldsymbol{j}\cdot\nabla\alpha(\boldsymbol{z})n^{-1}} \right)
\end{aligned} \tag{B.9}
$$

for some $\boldsymbol{z}$ on the line connecting $\boldsymbol{t}$ and $\boldsymbol{t}+\boldsymbol{j}/n$. To show $1 - n^{-1/n}$ is asymptotically equal to $n^{-1}\log(n)$ we can apply L'Hopital's rule, taking derivatives w.r.t. n,

$$
\begin{aligned}
\frac{1 - n^{-1/n}}{n^{-1}\log(n)} &= \frac{\exp(-n^{-1}\log(n))}{n^{-1}\log(n)} \\
&\overset{L'H.R.}{\Rightarrow} \frac{\exp(-n^{-1}\log(n))\,(-n^{-2} + \log(n)n^{-2})}{n^{-2} - \log(n)n^{-2}} \\
&\asymp \frac{\log(n)n^{-2-1/n}}{\log(n)n^{-2}}
\end{aligned}
$$

$$\to \quad \frac{\log(n)n^{-2}}{\log(n)n^{-2}} = 1. \tag{B.10}$$

So for e_3, applying (B.10), (B.8) and (15),

$$\begin{aligned}
|e_3| &\le n^{\tilde{\alpha}(\boldsymbol{t})-\alpha(\boldsymbol{t})} \sum_{i=0}^{x}\sum_{j=0}^{x}\binom{x}{i}\binom{x}{j}\left|\left(f_{\boldsymbol{t}+j\boldsymbol{h}/n}(\boldsymbol{h}) - f_{\boldsymbol{t}}(\boldsymbol{h})\right)|\boldsymbol{u}+(i-j)\boldsymbol{h}|^{\alpha(\boldsymbol{t})}\right| \\
&\le n^{-1}n^{O\left(n^{-1}\right)} \sum_{i=0}^{x}\sum_{j=0}^{x}\binom{x}{i}\binom{x}{j} j|\boldsymbol{h}|\mathbf{c}_{15}|\boldsymbol{u}+(i-j)\boldsymbol{h}|^{\alpha(\boldsymbol{t})}.
\end{aligned}$$

And so $|e_3| = O(n^{-1})$ uniformly.

Notice that $e_4 = 0$ if $\alpha(\boldsymbol{t})$ is constant. Then if we apply (B.9) and (B.10),

$$\begin{aligned}
|e_4| &= n^{\tilde{\alpha}(\boldsymbol{t})} \sum_{i=0}^{x}\sum_{j=0}^{x}(-1)^{i+j}\binom{x}{i}\binom{x}{j} f_{\boldsymbol{t}+i\boldsymbol{h}/n}(\boldsymbol{h}) \\
&\quad \cdot \left||\boldsymbol{u}/n+(i-j)\boldsymbol{h}/n|^{\alpha(\boldsymbol{t}+i\boldsymbol{h}/n)} - |\boldsymbol{u}/n+(i-j)\boldsymbol{h}/n|^{\alpha(\boldsymbol{t})}\right| \\
&= O\left(n^{-1+O(n^{-1})}\log(n)\right) \\
&= O\left(n^{-1}\log(n)\right).
\end{aligned}$$

Lastly, since $\alpha(\boldsymbol{t})$ is continuously differentiable and applying (B.10),

$$n^{\tilde{\alpha}(\boldsymbol{t})-\alpha(\boldsymbol{t}+i\boldsymbol{h}/n)-\gamma(\boldsymbol{t}+i\boldsymbol{h}/n)} = n^{-\gamma(\boldsymbol{t})}n^{O(n^{-1})} \to n^{-\gamma(\boldsymbol{t})}.$$

Therefore

$$\begin{aligned}
&|e_5| \\
&= n^{\tilde{\alpha}(\boldsymbol{t})}\left|\sum_{i=0}^{x}\sum_{j=0}^{x}(-1)^{i+j}\binom{x}{i}\binom{x}{j} O\left(|\boldsymbol{u}/n+(i-j)\boldsymbol{h}/n|^{\alpha(\boldsymbol{t}+i\boldsymbol{h}/n)+\gamma(\boldsymbol{t}+i\boldsymbol{h}/n)}\right)\right| \\
&= n^{\tilde{\alpha}(\boldsymbol{t})}\left|\sum_{i=0}^{x}\sum_{j=0}^{x} O\left(n^{-\alpha(\boldsymbol{t}+i\boldsymbol{h}/n)-\gamma(\boldsymbol{t}+i\boldsymbol{h}/n)}\right)\right| \\
&= O(n^{-\gamma(\boldsymbol{t})}).
\end{aligned}$$

Now we will prove (ii). Notice that for $|\boldsymbol{t}-\boldsymbol{s}| > (2|\boldsymbol{h}|+1)/n$,

$$\begin{aligned}
&E\Delta_{\boldsymbol{h}/n}(Y(\boldsymbol{t})-\mu(\boldsymbol{t}))\Delta_{\boldsymbol{u}/n}(Y(\boldsymbol{s})-\mu(\boldsymbol{s})) \\
&= \int_0^{1/n}\int_0^{1/n} \partial_{\boldsymbol{h}}^{(1,1)} C(\boldsymbol{t}+(1-\phi)\boldsymbol{h}, \boldsymbol{s}+(1-\eta)\boldsymbol{h})d\eta d\phi.
\end{aligned}$$

Therefore if $|\boldsymbol{t}-\boldsymbol{s}| > (x|\boldsymbol{h}|+1)/n$ and if we let $\boldsymbol{u}_i = \boldsymbol{h}$

$$\begin{aligned}
& n^{\alpha} E \Delta_{\boldsymbol{h}}^{x}(Y(\boldsymbol{t})-\mu(\boldsymbol{t}))\Delta_{\boldsymbol{h}}^{x}(Y(\boldsymbol{s})-\mu(\boldsymbol{s})) \\
= \;& n^{\alpha} E(\Delta_{\boldsymbol{u}_1/n}\cdots\Delta_{\boldsymbol{u}_x/n}Y(\boldsymbol{t})\Delta_{\boldsymbol{u}_1/n}\cdots\Delta_{\boldsymbol{u}_x/n}Y(\boldsymbol{s})) \\
= \;& n^{\alpha}\int_0^{1/n}\cdots\int_0^{1/n}\partial_{\boldsymbol{u}_1}^{(1,1)}\cdots\partial_{\boldsymbol{u}_x}^{(1,1)}C(\boldsymbol{t}+\sum_{i=1}^{x}(1-\phi_i)\boldsymbol{u}_i,\boldsymbol{s}+\sum_{i=1}^{x}(1-\eta_i)\boldsymbol{u}_i) \\
& \qquad \cdot d\phi_1 d\eta_1\cdots d\phi_x d\eta_x \\
* \;\leq \;& n^{\alpha(\boldsymbol{t})-2x}\mathbf{c}_{16}|\boldsymbol{t}-\boldsymbol{s}|^{\alpha(\boldsymbol{t})-2x}.
\end{aligned}$$

For some $\mathbf{c}_{16} > 0$. To show $*$, since $C(\boldsymbol{t},\boldsymbol{s})$ is $C^{(x,x)}$ away from 0 by Assumption 2.3(ii),

$$\begin{aligned}
& \int_0^{1/n}\cdots\int_0^{1/n}\partial_{\boldsymbol{u}_1}^{(1,1)}\cdots\partial_{\boldsymbol{u}_x}^{(1,1)}C(\boldsymbol{t}+\sum_{i=1}^{x}(1-\phi_i)\boldsymbol{u}_i,\boldsymbol{s}+\sum_{i=1}^{x}(1-\eta_i)\boldsymbol{u}_i) \\
& \qquad d\phi_1 d\eta_1\cdots d\phi_x d\eta_x \\
= \;& n^{-2x}\partial_{\boldsymbol{u}_1}^{(1,1)}\cdots\partial_{\boldsymbol{u}_x}^{(1,1)}v(\boldsymbol{z}_1,\boldsymbol{z}_2)
\end{aligned}$$

for some $\boldsymbol{z}_1$ in the integration region for $\boldsymbol{t}$ and $\boldsymbol{z}_2$ in the integration region for $\boldsymbol{s}$. Then by Assumption 2.3(ii), $\partial_{\boldsymbol{u}_1}^{(1,1)}\cdots\partial_{\boldsymbol{u}_x}^{(1,1)}R(\boldsymbol{z}_1,\boldsymbol{z}_2) \leq \mathbf{c}_2|\boldsymbol{t}-\boldsymbol{s}|^{\alpha-2x}$. Since $[0,1]^d$ is closed and $\partial_{\boldsymbol{u}_1}^{(1,1)}\cdots\partial_{\boldsymbol{u}_x}^{(1,1)}R(\boldsymbol{t},\boldsymbol{s})$ is continuous, we can apply the extreme value theorem and conclude that $\mathbf{c}_{16}$ is uniform over $[0,1]^d$. $\square$

Proof of Theorem 3.3

Proof. From Taylor's theorem and Assumptions 3.1(i) and (ii),

$$\begin{aligned}
E(W_n(\boldsymbol{t}_i)^2) &= C_n(\boldsymbol{t}_i,\boldsymbol{t}_i)+\mu(\boldsymbol{t}_i)^2 \\
&= g(\boldsymbol{t}_0)+(\boldsymbol{t}_i-\boldsymbol{t}_0)'\nabla g(\boldsymbol{t}_0)+(\boldsymbol{t}_i-\boldsymbol{t}_0)'H_g(\boldsymbol{t}_0)(\boldsymbol{t}_i-\boldsymbol{t}_0) \\
&\quad +o(\|\boldsymbol{t}_i-\boldsymbol{t}_0\|^2)+O(n^{-\rho(\boldsymbol{t}_i)})+O(n^{-2\delta(\boldsymbol{t}_0)})
\end{aligned} \tag{B.11}$$

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

$$M_0 = \begin{pmatrix} S \\ S_{1^1} \\ S_{2^1} \\ \vdots \\ S_{d^1} \end{pmatrix},\ M_i = \begin{pmatrix} S_{i^1} \\ S_{i^1 1^1} \\ S_{i^1 2^1} \\ \vdots \\ S_{i^1 d^1} \end{pmatrix},\ N_{i,j} = \begin{pmatrix} S_{i^1 j^1} \\ S_{i^1 j^1 1^1} \\ S_{i^1 j^1 2^1} \\ \vdots \\ S_{i^1 j^1 d^1} \end{pmatrix},\ \mathcal{N}_{i,j} = \begin{pmatrix} \kappa_{i^1 j^1} \\ \kappa_{i^1 j^1 1^1} \\ \kappa_{i^1 j^1 2^1} \\ \vdots \\ \kappa_{i^1 j^1 d^1} \end{pmatrix}.$$

It then follows from (B.11) that

$$
\begin{aligned}
X'KE(W^2) &= g(\boldsymbol{t}_0)M_0 + \sum_{i=1}^{d} \frac{\partial}{\partial t_i} g(\boldsymbol{t}_0)M_i && \text{(B.12)}\\
&+ \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial}{\partial t_i}\frac{\partial}{\partial t_j} g(\boldsymbol{t}_0)N_{i,j} && \text{(B.13)}\\
&+ \begin{pmatrix} \sum_{\boldsymbol{t}_i} k_{\boldsymbol{t}_i}\left\{o(\|\boldsymbol{t}_i - \boldsymbol{t}_0\|^2) + O(n^{-\rho(\boldsymbol{t}_i)}) + O(n^{-2\delta(\boldsymbol{t}_i)})\right\} \\ \sum_{\boldsymbol{t}_i} k_{\boldsymbol{t}_i} x_{\boldsymbol{i}1}\left\{o(\|\boldsymbol{t}_i - \boldsymbol{t}_0\|^2) + O(n^{-\rho(\boldsymbol{t}_i)}) + O(n^{-2\delta(\boldsymbol{t}_i)})\right\} \\ \vdots \\ \sum_{\boldsymbol{t}_i} k_{\boldsymbol{t}_i} x_{\boldsymbol{i}d}\left\{o(\|\boldsymbol{t}_i - \boldsymbol{t}_0\|^2) + O(n^{-\rho(\boldsymbol{t}_i)}) + O(n^{-2\delta(\boldsymbol{t}_i)})\right\} \end{pmatrix}. && \text{(B.14)}
\end{aligned}
$$

Since $(X'KX)^{-1}X'KX = I_{d+1}$, with I_{d+1} the $d+1 \times d+1$ identity matrix, and M_ℓ, $0 \le \ell \le d$, are the columns of $X'KX$, we have

$$
(X'KX)^{-1}M_\ell = e_{\ell+1},\ \ 0 \le \ell \le d, \tag{B.15}
$$

where e_ℓ is a column vector with a 1 in the ℓ^{th} row and zeros elsewhere. Thus, combining (B.12) and (B.15),

$$
(X'KX)^{-1}\left(g(\boldsymbol{t}_0)M_0 + \sum_{i=1}^{d} \frac{\partial}{\partial t_i} g(\boldsymbol{t}_0)M_i\right) = \beta(\boldsymbol{t}_0). \tag{B.16}
$$

By lemma B.1, if we let $D = diag(1, b, \ldots, b)$,

$$
\begin{aligned}
(X'KX)^{-1} &= \left[n^d b^d D\left(\mathcal{K} + O(n^{-1}b^{-1})\right)D\right]^{-1}\\
&= n^{-d}b^{-d}D^{-1}\left(\mathcal{K}^{-1} + O(n^{-1}b^{-1})\right)D^{-1}. && \text{(B.17)}
\end{aligned}
$$

Also,

$$
\begin{aligned}
&\sum_{i,j=1}^{d} \frac{\partial}{\partial t_i}\frac{\partial}{\partial t_j} g(\boldsymbol{t}_0)N_{i,j}\\
&= n^d b^{d+2} D \sum_{i,j=1}^{d} \frac{\partial}{\partial t_i}\frac{\partial}{\partial t_j} g(\boldsymbol{t}_0)\mathcal{N}_{i,j} + (nb)^{-1}O\begin{pmatrix} n^d b^{d+2} \\ n^d b^{d+3} \\ n^d b^{d+3} \\ \vdots \\ n^d b^{d+3} \end{pmatrix}. && \text{(B.18)}
\end{aligned}
$$

So combining (B.17) and (B.18),

$$
\begin{aligned}
& (X'KX)^{-1}\sum_{i,j=1}^{d}\frac{\partial}{\partial t_i}\frac{\partial}{\partial t_j}g(\boldsymbol{t}_0)N_{i,j} \\
= \;& b^2 D\sum_{i,j=1}^{d}\frac{\partial}{\partial t_i}\frac{\partial}{\partial t_j}g(\boldsymbol{t}_0)\mathcal{N}_{i,j}+O(n^{-1}b)D^{-1}\mathbf{1}. \qquad \text{(B.19)}
\end{aligned}
$$

Since $\rho(\boldsymbol{t})$ is continuously differentiable, $|\rho(\boldsymbol{t}_i)-\rho(\boldsymbol{t}_0)|\le \mathbf{c}_{17}|\boldsymbol{t}_i-\boldsymbol{t}_0|$ uniformly for some $\mathbf{c}_{17}>0$, and hence $n^{-\rho(\boldsymbol{t}_i)}\le n^{-\rho(\boldsymbol{t}_0)}n^{\mathbf{c}_{17}b}$. Then $n^{\mathbf{c}_{17}b}=\exp\{\mathbf{c}_{17}b\log(n)\}\to 1$ since $b=o(n^{-1/3})$. So $n^{-\rho(\boldsymbol{t}_i)}=O(n^{-\rho(\boldsymbol{t}_0)})$. Similarly, $n^{-2\delta(\boldsymbol{t}_i)}=O(n^{-2\delta(\boldsymbol{t}_0)})$. Therefore if we combine (B.14) and (B.17),

$$
\begin{aligned}
& (X'KX)^{-1}\begin{pmatrix}
\sum_{\boldsymbol{t}_i}k_{\boldsymbol{t}_i}\left\{o(\|\boldsymbol{t}_i-\boldsymbol{t}_0\|^2)+O(n^{-\rho(\boldsymbol{t}_i)})+O(n^{-2\delta(\boldsymbol{t}_i)})\right\}\\
\sum_{\boldsymbol{t}_i}k_{\boldsymbol{t}_i}x_{\boldsymbol{i}1}\left\{o(\|\boldsymbol{t}_i-\boldsymbol{t}_0\|^2)+O(n^{-\rho(\boldsymbol{t}_i)})+O(n^{-2\delta(\boldsymbol{t}_i)})\right\}\\
\vdots\\
\sum_{\boldsymbol{t}_i}k_{\boldsymbol{t}_i}x_{\boldsymbol{i}d}\left\{o(\|\boldsymbol{t}_i-\boldsymbol{t}_0\|^2)+O(n^{-\rho(\boldsymbol{t}_i)})+O(n^{-2\delta(\boldsymbol{t}_i)})\right\}
\end{pmatrix}\\
= \;& \begin{pmatrix}
o(b^2)+O(n^{-\rho(\boldsymbol{t}_0)})+O(n^{-2\delta(\boldsymbol{t}_0)})\\
o(b)+O(n^{-\rho(\boldsymbol{t}_0)}b^{-1})+O(n^{-2\delta(\boldsymbol{t}_0)}b^{-1})\\
o(b)+O(n^{-\rho(\boldsymbol{t}_0)}b^{-1})+O(n^{-2\delta(\boldsymbol{t}_0)}b^{-1})
\end{pmatrix}. \qquad \text{(B.20)}
\end{aligned}
$$

Thus, we obtain (3.1) by combining (B.16), (B.19) and (B.20).

□

Proof of Theorem 3.4

Proof. First we will prove the result for $x=2$. By (B.17), for any vector $\boldsymbol{v}$ of length n^d, we can write

$$
\begin{aligned}
& [(X'KX)^{-1}X'K\boldsymbol{v}]_1\\
= \;& (nb)^{-d}\sum_i k\left(\frac{\boldsymbol{t}_i-\boldsymbol{t}_0}{b}\right)\left\{\mathcal{K}_{1,1}^{-1}+\sum_{j=1}^{d}\mathcal{K}_{1,j+1}^{-1}\left(\frac{t_{ij}-t_{0j}}{b}\right)+O((nb)^{-1})\right\}v_i\\
=: \;& (nb)^{-d}\sum_i \tilde{k}\left(\frac{\boldsymbol{t}_i-\boldsymbol{t}_0}{b}\right)v_i.
\end{aligned}
$$

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

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

and $\widetilde{W}^2$ denote the vector with elements $\widetilde{W}_n(\boldsymbol{t}_i)^2$ and let $\tilde{\beta}(\boldsymbol{t}_0)$ be defined with $W_n(\boldsymbol{t}_i)$

replaced by $\widetilde{W}_n(\boldsymbol{t}_i)$. Then, we can write

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

If we apply Corollary B.4,

$$\mathbb{E}\left(\tilde{\beta}_0(\boldsymbol{t}_0) - \mathbb{E}(\tilde{\beta}_0(\boldsymbol{t}_0))\right)^2 = 2(nb)^{-2d} \sum_{\boldsymbol{t}_i,\boldsymbol{t}_i} \tilde{k}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right) \tilde{k}\left(\frac{\boldsymbol{t}_j - \boldsymbol{t}_0}{b}\right) C_n^2(\boldsymbol{t}_i, \boldsymbol{t}_j). \quad \text{(B.21)}$$

Then if $\tilde{\boldsymbol{t}}_0$ is the closest gridpoint to $\boldsymbol{t}_0$,

$$\begin{aligned}
& (nb)^{-d} \Bigg| \sum_{\boldsymbol{t}_i,\boldsymbol{t}_i} \Bigg(\tilde{k}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b} + \frac{\boldsymbol{t}_0 - \tilde{\boldsymbol{t}}_0}{b}\right) \tilde{k}\left(\frac{\boldsymbol{t}_j - \boldsymbol{t}_0}{b} + \frac{\boldsymbol{t}_0 - \tilde{\boldsymbol{t}}_0}{b}\right) \\
&\qquad - \tilde{k}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right) \tilde{k}\left(\frac{\boldsymbol{t}_j - \boldsymbol{t}_0}{b}\right) \Bigg) C_n^2(\boldsymbol{t}_i, \boldsymbol{t}_j) \Bigg| \\
\leq\; & (nb)^{-d} \Bigg| \sum_{\boldsymbol{t}_i,\boldsymbol{t}_i} \Bigg(\tilde{k}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b} + \frac{\boldsymbol{t}_0 - \tilde{\boldsymbol{t}}_0}{b}\right) \tilde{k}\left(\frac{\boldsymbol{t}_j - \boldsymbol{t}_0}{b} + \frac{\boldsymbol{t}_0 - \tilde{\boldsymbol{t}}_0}{b}\right) \\
&\qquad - \tilde{k}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b} + \frac{\boldsymbol{t}_0 - \tilde{\boldsymbol{t}}_0}{b}\right) \tilde{k}\left(\frac{\boldsymbol{t}_j - \boldsymbol{t}_0}{b}\right) \Bigg) C_n^2(\boldsymbol{t}_i, \boldsymbol{t}_j) \Bigg| \\
& + (nb)^{-d} \Bigg| \sum_{\boldsymbol{t}_i,\boldsymbol{t}_i} \Bigg(\tilde{k}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b} + \frac{\boldsymbol{t}_0 - \tilde{\boldsymbol{t}}_0}{b}\right) \tilde{k}\left(\frac{\boldsymbol{t}_j - \boldsymbol{t}_0}{b}\right) \\
&\qquad - \tilde{k}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right) \tilde{k}\left(\frac{\boldsymbol{t}_j - \boldsymbol{t}_0}{b}\right) \Bigg) C_n^2(\boldsymbol{t}_i, \boldsymbol{t}_j) \Bigg| \\
\leq\; & (nb)^{-d} \sum_{\boldsymbol{t}_i,\boldsymbol{t}_i} \mathbf{c}_{18} \frac{1}{nb} \left| \tilde{k}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b} + \frac{\boldsymbol{t}_0 - \tilde{\boldsymbol{t}}_0}{b}\right) \right| C_n^2(\boldsymbol{t}_i, \boldsymbol{t}_j) \\
& + (nb)^{-d} \sum_{\boldsymbol{t}_i,\boldsymbol{t}_i} \mathbf{c}_{18} \frac{1}{nb} \left| \tilde{k}\left(\frac{\boldsymbol{t}_j - \boldsymbol{t}_0}{b}\right) \right| C_n^2(\boldsymbol{t}_i, \boldsymbol{t}_j) \qquad \text{(B.22)}
\end{aligned}$$

for some $\mathbf{c}_{18} > 0$ since k is continously differentiable. It follows from comments below that (B.22) is bounded uniformly by $\mathbf{c}_{19}(nb)^{-1}$ for some $\mathbf{c}_{19} > 0$.

Relabel the points $\boldsymbol{t}_i$ as $\tilde{\boldsymbol{t}}_0 + \boldsymbol{i}/n, \boldsymbol{i} \in \mathbb{Z}^2$ and denote by $S_{\boldsymbol{i}}$ the square centered at $\boldsymbol{i}/(nb)$ with width $1/(nb)$. Then for a fixed positive m, write

$$\begin{aligned}
& (nb)^{-d} \sum_{\boldsymbol{i},\boldsymbol{j}} \tilde{k}\left(\frac{\boldsymbol{i}}{nb}\right) \tilde{k}\left(\frac{\boldsymbol{j}}{nb}\right) C_n^2(\tilde{\boldsymbol{t}}_0 + \boldsymbol{i}/n, \tilde{\boldsymbol{t}}_0 + \boldsymbol{j}/n) - \int_{I(\boldsymbol{t}_0,b)} \bar{k}^2(\mathbf{z}) d\mathbf{z} \sum_{\boldsymbol{j} \in \mathbb{Z}^2} g_2^2(\boldsymbol{t}_0, \boldsymbol{j}) \\
=\; & \epsilon_{n,1} + \epsilon_{n,2} + \epsilon_{n,3} + \epsilon_{n,4} + \epsilon_{n,5},
\end{aligned}$$

where

$$
\begin{aligned}
\epsilon_{n,1} &= (nb)^{-d}\left(\sum_{\boldsymbol{i},\boldsymbol{j}} - \sum_{\boldsymbol{i}}\sum_{\boldsymbol{j}:|\boldsymbol{j}-\boldsymbol{i}|\le m}\right)\tilde{k}\left(\frac{\boldsymbol{i}}{nb}\right)\tilde{k}\left(\frac{\boldsymbol{j}}{nb}\right)C_n^2(\tilde{\mathbf{t}}_0+\boldsymbol{i}/n,\tilde{\mathbf{t}}_0+\boldsymbol{j}/n),\\
\epsilon_{n,2} &= (nb)^{-d}\Bigg(\sum_{\boldsymbol{i}}\sum_{|\boldsymbol{j}|\le m}\tilde{k}\left(\frac{\boldsymbol{i}}{nb}\right)\tilde{k}\left(\frac{\boldsymbol{i}+\boldsymbol{j}}{nb}\right)C_n^2(\tilde{\mathbf{t}}_0+\boldsymbol{i}/n,\tilde{\mathbf{t}}_0+(\boldsymbol{i}+\boldsymbol{j})/n)\\
&\qquad\qquad - \sum_{\boldsymbol{i}}\tilde{k}^2\left(\frac{\boldsymbol{i}}{nb}\right)\sum_{|\boldsymbol{j}|\le m}C_n^2(\tilde{\mathbf{t}}_0+\boldsymbol{i}/n,\tilde{\mathbf{t}}_0+(\boldsymbol{i}+\boldsymbol{j})/n)\Bigg),\\
\epsilon_{n,3} &= (nb)^{-d}\sum_{\boldsymbol{i}}\tilde{k}^2\left(\frac{\boldsymbol{i}}{nb}\right)\sum_{|\boldsymbol{j}|\le m}C_n^2(\tilde{\mathbf{t}}_0+\boldsymbol{i}/n,\tilde{\mathbf{t}}_0+(\boldsymbol{i}+\boldsymbol{j})/n)\\
&\qquad - \sum_{\boldsymbol{i}}\int\limits_{S_{\boldsymbol{i}}}\bar{k}^2(\mathbf{z})d\boldsymbol{z}\sum_{|\boldsymbol{j}|\le m}C_n^2(\tilde{\mathbf{t}}_0+\boldsymbol{i}/n,\tilde{\mathbf{t}}_0+(\boldsymbol{i}+\boldsymbol{j})/n),\\
\epsilon_{n,4} &= \sum_{\boldsymbol{i}}\int\limits_{S_{\boldsymbol{i}}}\bar{k}^2(\mathbf{z})d\boldsymbol{z}\sum_{|\boldsymbol{j}|\le m}C_n^2(\tilde{\mathbf{t}}_0+\boldsymbol{i}/n,\tilde{\mathbf{t}}_0+(\boldsymbol{i}+\boldsymbol{j})/n)\\
&\qquad - \int\limits_{I(\boldsymbol{t}_0,b)}\bar{k}^2(\boldsymbol{z})d\boldsymbol{z}\sum_{|\boldsymbol{j}|\le m}g(\boldsymbol{t}_0,\boldsymbol{j})^2,\\
\epsilon_{n,5} &= \int\limits_{I(\boldsymbol{t}_0,b)}\bar{k}^2(\boldsymbol{z})d\boldsymbol{z}\sum_{|\boldsymbol{j}|\le m}g(\boldsymbol{t}_0,\boldsymbol{j})^2 - \int\limits_{I(\boldsymbol{t}_0,b)}\bar{k}^2(\boldsymbol{z})d\boldsymbol{z}g_2^2(\boldsymbol{t}_0,\boldsymbol{j}).
\end{aligned}
$$

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

$$\lim_{m\to\infty}\limsup_{n\to\infty}\epsilon_{n,i} = 0. \tag{B.23}$$

First,

$$\epsilon_{n,1} = (nb)^{-d}\sum_{\boldsymbol{i}}\sum_{|\boldsymbol{j}|>m}\tilde{k}\left(\frac{\boldsymbol{i}}{nb}\right)\tilde{k}\left(\frac{\boldsymbol{i}+\boldsymbol{j}}{nb}\right)C_n^2(\tilde{\mathbf{t}}_0+\boldsymbol{i}/n,\tilde{\mathbf{t}}_0+(\boldsymbol{i}+\boldsymbol{j})/n)$$

Fix $m > \tau$. By Assumption 3.1(iv), for all $\boldsymbol{j}$ with $|\boldsymbol{j}| > m$,

$$\sup_n\sup_{\boldsymbol{j}:|\boldsymbol{j}|>m}\frac{C_n^2(\tilde{\mathbf{t}}_0+\boldsymbol{i}/n,\tilde{\mathbf{t}}_0+(\boldsymbol{i}+\boldsymbol{j})/n)}{|\boldsymbol{j}|^{-2\psi(\boldsymbol{t}_0)}} \le \mathbf{c}_6^2. \tag{B.24}$$

Thus,

$$\epsilon_{n,1} \le \mathbf{c}_6^2(nb)^{-d}\sum_{\boldsymbol{i}}\tilde{k}\left(\frac{\boldsymbol{i}}{nb}\right)\sum_{|\boldsymbol{j}|>m}\tilde{k}\left(\frac{\boldsymbol{i}+\boldsymbol{j}}{nb}\right)|\boldsymbol{j}|^{-2\psi(\boldsymbol{t}_0)}.$$

Since k is bounded and $-2\psi(\boldsymbol{t}_0) \leq -1$, (B.23) holds for $i = 1$.

Next, since k has bounded first derivative,

$$\left|\tilde{k}\left(\frac{\boldsymbol{i}+\boldsymbol{j}}{nb}\right) - \tilde{k}\left(\frac{\boldsymbol{i}}{nb}\right)\right| \leq \mathbf{c}_{20}|\boldsymbol{j}|(nb)^{-1} \tag{B.25}$$

for some $\mathbf{c}_{20} > 0$. Combining (B.25) with Assumption 3.1,

$$\begin{aligned} & |\epsilon_{n,2}| \\ \leq \;& (nb)^{-d}\sum_{\boldsymbol{i}}\sum_{|\boldsymbol{j}|\leq m}\tilde{k}\left(\frac{\boldsymbol{i}}{nb}\right)\left|\tilde{k}\left(\frac{\boldsymbol{i}+\boldsymbol{j}}{nb}\right) - \tilde{k}\left(\frac{\boldsymbol{i}}{nb}\right)\right| C_n^2(\tilde{\mathbf{t}}_0 + \boldsymbol{i}/n, \tilde{\mathbf{t}}_0 + (\boldsymbol{i}+\boldsymbol{j})/n) \\ \leq \;& \mathbf{c}_{21}(nb)^{-d-1}\sum_{\boldsymbol{i}}\tilde{k}\left(\frac{\boldsymbol{i}}{nb}\right)\left(\sum_{|\boldsymbol{j}|\leq\tau}\mathbf{c}_5 + \sum_{\tau<|\boldsymbol{j}|\leq m}\mathbf{c}_6|\boldsymbol{j}|^{-2\psi(\boldsymbol{t}_0)}\right), \end{aligned}$$

for some $\mathbf{c}_{21} > 0$. From which (B.23) follows for $i = 2$.

By the mean-value theorem, write $\int_{S_{\boldsymbol{i}}} \bar{k}^2(\mathbf{z})d\boldsymbol{z} = (nb)^{-d}\bar{k}^2(\boldsymbol{z}_{\boldsymbol{i}})$ for some $\boldsymbol{z}_{\boldsymbol{i}} \in S_{\boldsymbol{i}}$. By the boundedness of k' and $C_n^2(\mathbf{s}, \mathbf{t})$, there exists some $\mathbf{c}_{22} > 0$ s.t.

$$\begin{aligned} |\epsilon_{n,3}| &= \left|(nb)^{-d}\sum_{\boldsymbol{i}}\left(\tilde{k}^2\left(\frac{\boldsymbol{i}}{nb}\right) - \bar{k}^2(\boldsymbol{z}_{\boldsymbol{i}})\right)\sum_{|\boldsymbol{j}|\leq m} C_n^2(\tilde{\mathbf{t}}_0 + \boldsymbol{i}/n, \tilde{\mathbf{t}}_0 + (\boldsymbol{i}+\boldsymbol{j})/n)\right| \\ &\leq \mathbf{c}_{22}m(nb)^{-d+1}, \end{aligned}$$

which shows that (B.23) holds for $i = 3$.

Assumption 3.1(ii) readily establishes (B.23) for $i = 4$.

By Fatou's Lemma, Assumption 3.1(ii), (B.24) and the definition of $A(\boldsymbol{t}_0, b)$,

$$A(\boldsymbol{t}_0, b) \leq \liminf_{n\to\infty} \sum_{\boldsymbol{j}\in\mathbb{Z}^d\cap C(0,nb)} C_n^2(\mathbf{t}_0 + \boldsymbol{i}/n, \mathbf{t}_0 + (\boldsymbol{i}+\boldsymbol{j})/n) < \infty.$$

This shows that (B.23) holds for $i = 5$ and completes the proof for $x = 2$.

For general $x \geq 2$,

$$\begin{aligned} & n^{xd/2}b^{xd/2}E[\tilde{\beta}_0(\boldsymbol{t}_0; n, b) - E(\tilde{\beta}_0(\boldsymbol{t}_0; n, b))]^x \\ = \;& (nb)^{-xd/2}\sum_{\boldsymbol{t}_{i_1},\ldots,\boldsymbol{t}_{i_x}} E\prod_{j=1}^{x}\tilde{k}_{\boldsymbol{t}_{ij}}[\widetilde{W}(\boldsymbol{t}_{ij}) - E\widetilde{W}(\boldsymbol{t}_{ij})]^2. \end{aligned} \tag{B.26}$$

Now let $\delta_{\boldsymbol{t}_i\boldsymbol{t}_j} = \tilde{k}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right)^{1/2}\tilde{k}\left(\frac{\boldsymbol{t}_j - \boldsymbol{t}_0}{b}\right)^{1/2} C_n(\boldsymbol{t}_i, \boldsymbol{t}_j)$. Then from the Gaussian moment theorem, (B.26) becomes

$$(nb)^{-xd/2}\sum_{\boldsymbol{t}_{i_1},\ldots,\boldsymbol{t}_{i_x}}\sum_{\mathbf{S}}\delta_{\boldsymbol{t}_{s_1}\boldsymbol{t}_{s_2}}\delta_{\boldsymbol{t}_{s_3}\boldsymbol{t}_{s_4}}\cdots\delta_{\boldsymbol{t}_{s_{2x-1}}\boldsymbol{t}_{s_{2x}}},$$

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

To explain the idea of the general proof, we first consider the cases $x=3$ and 4. For $x=3$, we have

$$(nb)^{3d/2}E\left(\tilde{\beta}_0(\boldsymbol{t}_0;n,b)-E\tilde{\beta}_0(\boldsymbol{t}_0;n,b)\right)^3=(nb)^{-3d/2}\sum_{\boldsymbol{t}_{i_1},\boldsymbol{t}_{i_2},\boldsymbol{t}_{i_3}}\sum_{\mathbf{S}}\delta_{\boldsymbol{t}_{s_1}\boldsymbol{t}_{s_2}}\delta_{\boldsymbol{t}_{s_3},\boldsymbol{t}_{s_4}}\delta_{\boldsymbol{t}_{s_5}\boldsymbol{t}_{s_6}},$$

where $\mathbf{S}$ is the set of all possible ways to make three pairs, $\{(s_1,s_2),(s_3,s_4),(s_5,s_6)\}$, with s_j chosen from $\mathbf{I}^3=\{i_1,i_1,i_2,i_2,i_3,i_3\}$ without replacement and each pair must be of different indices. Observe that for any given indices i_1,i_2,i_3, the number of all possible pairings in $\mathbf{S}$ 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, by lemma B.5

$$\begin{aligned}
&\left|(nb)^{-3d/2}\sum_{\boldsymbol{t}_{i_1},\boldsymbol{t}_{i_2},\boldsymbol{t}_{i_3}}\sum_{\mathbf{S}}\delta_{\boldsymbol{t}_{s_1}\boldsymbol{t}_{s_2}}\delta_{\boldsymbol{t}_{s_3},\boldsymbol{t}_{s_4}}\delta_{\boldsymbol{t}_{s_5}\boldsymbol{t}_{s_6}}\right|\\
=&\left|(nb)^{-3d/2}8\sum_{\boldsymbol{t}_{i_1},\boldsymbol{t}_{i_2},\boldsymbol{t}_{i_3}}\delta_{\boldsymbol{t}_{i_1}\boldsymbol{t}_{i_2}}\delta_{\boldsymbol{t}_{i_1},\boldsymbol{t}_{i_3}}\delta_{\boldsymbol{t}_{i_2}\boldsymbol{t}_{i_3}}\right|\\
\leq&\ (nb)^{-3d/2}8\sum_{\boldsymbol{t}_{i_1},\boldsymbol{t}_{i_2}}|\delta_{\boldsymbol{t}_{i_1}\boldsymbol{t}_{i_2}}|\frac{1}{2}\left(\sum_{\boldsymbol{t}_{i_3}}|\delta_{\boldsymbol{t}_{i_1},\boldsymbol{t}_{i_3}}|^2+|\delta_{\boldsymbol{t}_{i_2},\boldsymbol{t}_{i_3}}|^2\right)\\
\leq&\ (nb)^{-3d/2}\mathbf{c}_{23}\sum_{\boldsymbol{t}_{i_1},\boldsymbol{t}_{i_2}}|\delta_{\boldsymbol{t}_{i_1}\boldsymbol{t}_{i_2}}|\\
\leq&\ (nb)^{-d/2}\mathbf{c}_{24}r_{nb}(\boldsymbol{t}_0)
\end{aligned}$$

for sufficiently large $\mathbf{c}_{23},\mathbf{c}_{24}>0$.

For $x=4$, we have

$$\begin{aligned}
&(nb)^{2d}E\left(\tilde{\beta}_0(\boldsymbol{t}_0;n,b)-E(\tilde{\beta}_0(\boldsymbol{t}_0;n,b)\right)^4\\
=&\ (nb)^{-2d}\sum_{\boldsymbol{t}_{i_1},\boldsymbol{t}_{i_2},\boldsymbol{t}_{i_3},\boldsymbol{t}_{i_4}}\sum_{\mathbf{S}}\delta_{\boldsymbol{t}_{s_1}\boldsymbol{t}_{s_2}}\delta_{\boldsymbol{t}_{s_3}\boldsymbol{t}_{s_4}}\delta_{\boldsymbol{t}_{s_5}\boldsymbol{t}_{s_6}}\delta_{\boldsymbol{t}_{s_7}\boldsymbol{t}_{s_8}},
\end{aligned}\tag{B.27}$$

where $\mathbf{S}$ is the set of all possible ways to make 4 pairs, $\{(s_1, s_2), (s_3, s_4), (s_5, s_6), (s_7, s_8)\}$, with s_j chosen from $\mathbf{I}^4 = \{i_1, i_1, i_2, i_2, i_3, i_3, i_4, i_4\}$ without replacement and each pair must be of different indices. Only two configurations are possible:

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

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, (B.27) becomes

$$6(nb)^{-2d} \sum_{\boldsymbol{t}_{i_1}, \boldsymbol{t}_{i_2}, \boldsymbol{t}_{i_3}, \boldsymbol{t}_{i_4}} \delta_{\boldsymbol{t}_{i_1}\boldsymbol{t}_{i_2}} \delta_{\boldsymbol{t}_{i_2}\boldsymbol{t}_{i_3}} \delta_{\boldsymbol{t}_{i_3}\boldsymbol{t}_{i_4}} \delta_{\boldsymbol{t}_{i_4}\boldsymbol{t}_{i_1}} \tag{B.28}$$

$$+12(nb)^{-2d} \sum_{\boldsymbol{t}_{i_1}, \boldsymbol{t}_{i_2}} \delta^2_{\boldsymbol{t}_{i_1}\boldsymbol{t}_{i_2}} \sum_{\boldsymbol{t}_{i_3}, \boldsymbol{t}_{i_4}} \delta^2_{\boldsymbol{t}_{i_3}\boldsymbol{t}_{i_4}}. \tag{B.29}$$

For (B.28),

$$\begin{aligned} & (nb)^{-2d} \left| 6 \sum_{\boldsymbol{t}_{i_1}, \boldsymbol{t}_{i_2}, \boldsymbol{t}_{i_3}, \boldsymbol{t}_{i_4}} \delta_{\boldsymbol{t}_{i_1}\boldsymbol{t}_{i_2}} \delta_{\boldsymbol{t}_{i_2}\boldsymbol{t}_{i_3}} \delta_{\boldsymbol{t}_{i_3}\boldsymbol{t}_{i_4}} \delta_{\boldsymbol{t}_{i_4}\boldsymbol{t}_{i_1}} \right| \\ \leq \; & 6(nb)^{-2d}\mathbf{c}_5 \sum_{\boldsymbol{t}_{i_1}, \boldsymbol{t}_{i_2}, \boldsymbol{t}_{i_3}} |\delta_{\boldsymbol{t}_{i_1}\boldsymbol{t}_{i_2}} \delta_{\boldsymbol{t}_{i_2}\boldsymbol{t}_{i_3}}| \left(\sum_{\boldsymbol{t}_{i4}} |\delta_{\boldsymbol{t}_{i_3}\boldsymbol{t}_{i_4}}|^2 + |\delta_{\boldsymbol{t}_{i_4}\boldsymbol{t}_{i_1}}|^2 \right) \\ \leq \; & \mathbf{c}_{25}(nb)^{-d} r_{nb}(\boldsymbol{t})^2, \end{aligned} \tag{B.30}$$

for sufficiently large $\mathbf{c}_{25}$. Then (B.29) converges to $3A(\boldsymbol{t}_0, b)^2$ with error $o(1)$ from the proof when $x = 2$.

Now consider any general $x \geq 3$. Since each index i_j appears exactly twice, each pairing $\{(s_1, s_2), (s_3, s_4), \ldots, (s_{2x-1}, s_{2x})\}$ can be partitioned into a collection of subsets, "chains" of the form $\{(j_1, j_2), (j_2, j_3), \ldots, (j_q, j_1)\}$. For convenience, we say the chain $\{(j_1, j_2), (j_2, j_3), \ldots, (j_q, j_1)\}$ has length $q \geq 2$. Then similar to (25),

$$\begin{aligned} & (nb)^{-qd/2} \left| \sum_{\boldsymbol{t}_{i_1}, \ldots, \boldsymbol{t}_{i_q}} \delta_{\boldsymbol{t}_{i_1}\boldsymbol{t}_{i_2}} \cdots \delta_{\boldsymbol{t}_{i_q}\boldsymbol{t}_{i_1}} \right| \\ \leq \; & (nb)^{-qd/2} \sum_{\boldsymbol{t}_{i_1}} \sum_{\boldsymbol{t}_{i_2}, \ldots, \boldsymbol{t}_{i_{q-1}}} \left| \delta_{\boldsymbol{t}_{i_1}\boldsymbol{t}_{i_2}} \cdots \delta_{\boldsymbol{t}_{i_{q-2}}\boldsymbol{t}_{i_{q-1}}} \right| \left(\sum_{\boldsymbol{t}_{i_q}} \delta^2_{\boldsymbol{t}_{i_{q-1}}\boldsymbol{t}_{i_q}} + \delta^2_{\boldsymbol{t}_{i_q}\boldsymbol{t}_{i_1}} \right) \\ \leq \; & \mathbf{c}_{26} \left(\frac{r_{nb}(\boldsymbol{t})}{(nb)^{d/2}} \right)^{q-2}, \end{aligned} \tag{B.31}$$

for some $\mathbf{c}_{26} > 0$.

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 (26),

$$(nb)^{-xd/2} \sum_{\boldsymbol{t}_{i_1}, \boldsymbol{t}_{i_2}, \ldots, \boldsymbol{t}_{i_x}} \delta_{\boldsymbol{t}_{s_1}\boldsymbol{t}_{s_2}} \delta_{\boldsymbol{t}_{s_3}\boldsymbol{t}_{s_4}} \cdots \delta_{\boldsymbol{t}_{s_{2x-1}}\boldsymbol{t}_{s_{2x}}} \leq \mathbf{c}_{26}^m \left(\frac{r_{nb}(\boldsymbol{t}_0)}{(nb)^{d/2}}\right)^{\sum_{i=1}^m (\ell_i - 2)}. \tag{B.32}$$

If $m < x/2$, then $\ell_i > 2$ for some i and therefore $\sum_{i=1}^m (\ell_i - 2) > 0$ and (B.32) converges to 0. If $m = x/2$, then this partition will contain $x/2$ chains of length 2 and therefore

$$\begin{aligned}
& (nb)^{-xd/2} \sum_{\boldsymbol{t}_{i_1}, \boldsymbol{t}_{i_2}, \ldots, \boldsymbol{t}_{i_k}} \delta_{\boldsymbol{t}_{s_1}\boldsymbol{t}_{s_2}} \delta_{\boldsymbol{t}_{s_3}\boldsymbol{t}_{s_4}} \cdots \delta_{\boldsymbol{t}_{s_{2x-1}}\boldsymbol{t}_{s_{2x}}} \\
= \; & (nb)^{-xd/2} \Big(\sum_{\boldsymbol{t}_{i_1}, \boldsymbol{t}_{i_2}} \delta^2_{\boldsymbol{t}_{s_1}\boldsymbol{t}_{s_2}} \Big)^{x/2} \\
= \; & 2^{-x/2} A(\boldsymbol{t}_0, b)^{x/2} + o(1).
\end{aligned}$$

And the number of ways to obtain $x/2$ chains of length two from $\mathbf{I}^x$ is $(2(x-1)) \cdot (2(x-3)) \cdots 2 = (x-1)!!2^{x/2}$. So therefore if x is even, then

$$(nb)^{-xd/2} \sum_{\boldsymbol{t}_{i_1}, \boldsymbol{t}_{i_2}, \ldots, \boldsymbol{t}_{i_k}} \delta_{\boldsymbol{t}_{s_1}\boldsymbol{t}_{s_2}} \delta_{\boldsymbol{t}_{s_3}\boldsymbol{t}_{s_4}} \cdots \delta_{\boldsymbol{t}_{s_{2x-1}}\boldsymbol{t}_{s_{2x}}} = (x-1)!! A(\boldsymbol{t}_0, b)^{x/2} + o(1).$$

If x is odd then m can never be equal to $x/2$ and therefore

$$(nb)^{-xd/2} \sum_{\boldsymbol{t}_{i_1}, \boldsymbol{t}_{i_2}, \ldots, \boldsymbol{t}_{i_k}} \sum_{\mathbf{S}} \delta_{\boldsymbol{t}_{s_1}\boldsymbol{t}_{s_2}} \delta_{\boldsymbol{t}_{s_3}\boldsymbol{t}_{s_4}} \cdots \delta_{\boldsymbol{t}_{s_{2x-1}}\boldsymbol{t}_{s_{2x}}} = O\left(\frac{r_{nb}(\boldsymbol{t}_0)}{(nb)^{d/2}}\right).$$

□

Proof of Theorem 3.5

Proof.

$$\begin{aligned}
& A(\boldsymbol{t}_0, b)^{-1/2} (nb)^{d/2} \{\hat{\beta}_0(\boldsymbol{t}_0; n, b) - g(\boldsymbol{t}_0)\} \\
= \; & A(\boldsymbol{t}_0, b)^{-1/2} (nb)^{d/2} \left\{ \sum_{\boldsymbol{t}_i} \tilde{k}_{\boldsymbol{t}_i} W_n^2(\boldsymbol{t}_i) - g(\boldsymbol{t}_0) \right\} \\
= \; & A(\boldsymbol{t}_0, b)^{-1/2} (nb)^{d/2} \left\{ \sum_{\boldsymbol{t}_i} \tilde{k}_{\boldsymbol{t}_i} (W_n(\boldsymbol{t}_i) - \mu_n(\boldsymbol{t}_i) + \mu_n(\boldsymbol{t}_i))^2 - g(\boldsymbol{t}_0) \right\} \\
= \; & A(\boldsymbol{t}_0, b)^{-1/2} (nb)^{d/2} \sum_{\boldsymbol{t}_i} \tilde{k}_{\boldsymbol{t}_i} \Big\{ (W_n(\boldsymbol{t}_i) - \mu_n(\boldsymbol{t}_i))^2 - C_n(\boldsymbol{t}_i, \boldsymbol{t}_i)
\end{aligned} \tag{B.33}$$

$$
\left. \begin{aligned} &+E(W_n(\boldsymbol{t}_i))^2 - g(\boldsymbol{t}_0) \qquad (B.34) \\ &+2\mu_n(\boldsymbol{t}_i)(W_n(\boldsymbol{t}_i) - \mu_n(\boldsymbol{t}_i)) \end{aligned} \right\}. \qquad (B.35)
$$

For (B.33), from Theorem 3.4,

$$
\begin{aligned}
& A(\boldsymbol{t}_0, b)^{-1/2}(nb)^{d/2} E\left(\sum_{\boldsymbol{t}_i} \tilde{k}_{\boldsymbol{t}_i}\left\{(W_n(\boldsymbol{t}_i) - \mu_n(\boldsymbol{t}_i))^2 - C_n(\boldsymbol{t}_i, \boldsymbol{t}_i)\right\}\right)^x \\
= \quad & (x-1)!! + o(1) \\
\to \quad & 0
\end{aligned}
$$

for x even and $O\left(\frac{r_{nb}(\boldsymbol{t}_0)}{(nb)^{d/2}}\right) \to 0$ for x odd, which are the moments of the $N(0,1)$ distribution. So

$$
A(\boldsymbol{t}_0, b)^{-1/2}(nb)^{d/2} \sum_{\boldsymbol{t}_i} \tilde{k}_{\boldsymbol{t}_i}\left\{(W_n(\boldsymbol{t}_i) - \mu_n(\boldsymbol{t}_i))^2 - C_n(\boldsymbol{t}_i, \boldsymbol{t}_i)\right\} \stackrel{d}{=} N(0,1) + O_p(1)
$$

by Theorem B.6. Then for (B.34), from Theorem 3.3,

$$
\begin{aligned}
& A(\boldsymbol{t}_0, b)^{-1/2}(nb)^{d/2}\left(E\sum_{\boldsymbol{t}_i} \tilde{k}_{\boldsymbol{t}_i}(W_n(\boldsymbol{t}_i))^2 - g(\boldsymbol{t}_0)\right) \\
= \quad & \frac{1}{2A(\boldsymbol{t}_0, b)^{1/2}} n^{d/2} b^{d/2+2} \mathcal{K}\mathcal{M}\boldsymbol{g}^{(2)}(\boldsymbol{t}_0) + o(n^{d/2}b^{d/2+2}) + O(n^{d/2-\rho(\boldsymbol{t}_0)}b) \\
\to \quad & 0
\end{aligned}
$$

when $b = o(n^{-d/(d+4)\wedge(\rho(\boldsymbol{t}_0)-d/2)})$. And for (B.35), by assumption,

$$
W_n(\boldsymbol{t}_i) - \mu_n(\boldsymbol{t}_i) \sim N(0, C_n(\boldsymbol{t}_i, \boldsymbol{t}_i)).
$$

Therefore

$$
\begin{aligned}
& (nb)^{d/2} \sum_{\boldsymbol{t}_i} \tilde{k}_{\boldsymbol{t}_i} 2\mu_n(\boldsymbol{t}_i)(W_n(\boldsymbol{t}_i) - \mu_n(\boldsymbol{t}_i)) \\
\sim \quad & N\left(0, (nb)^d \sum_{\boldsymbol{t}_i, \boldsymbol{t}_j} 4\tilde{k}_{\boldsymbol{t}_i}\tilde{k}_{\boldsymbol{t}_j}\mu_n(\boldsymbol{t}_i)\mu_n(\boldsymbol{t}_j)C_n(\boldsymbol{t}_i, \boldsymbol{t}_j)\right)
\end{aligned}
$$

and $(nb)^d \sum_{\boldsymbol{t}_i, \boldsymbol{t}_j} 4\tilde{k}_{\boldsymbol{t}_i}\tilde{k}_{\boldsymbol{t}_j}\mu_n(\boldsymbol{t}_i)\mu_n(\boldsymbol{t}_j)C_n(\boldsymbol{t}_i, \boldsymbol{t}_j) = O(n^{-2\delta(\boldsymbol{t}_0)})r_{nb}(\boldsymbol{t})$.

Therefore

$$
A(\boldsymbol{t}_0, b)^{-1/2}(nb)^{d/2}\{\hat{\beta}_0(\boldsymbol{t}_0; n, b) - g(\boldsymbol{t}_0)\}
$$

$$
\begin{aligned}
&= Z + O\left(n^{d/2} b^{2+d/2}\right) + O\left(n^{-\rho(\boldsymbol{t}_0)+d/2} b^{d/2}\right) + O\left(n^{-2\delta(\boldsymbol{t}_0)+d/2} b^{d/2}\right) \\
&\quad + O_p\left(n^{-\delta(\boldsymbol{t}_0)+d/2} b^{d/2} \sqrt{r_{nb}(\boldsymbol{t})}\right) + o(1) \\
&\to Z,
\end{aligned}
$$

which implies that,

$$
\begin{aligned}
&\hat{\beta}_0(\boldsymbol{t}_0; n, b) \\
\overset{d}{=}\ & g(\boldsymbol{t}_0) + \frac{Z A(\boldsymbol{t}_0, b)^{1/2}}{(nb)^{d/2}} + \frac{1}{2} b^2 [\mathcal{K}\mathcal{M}\boldsymbol{g}^{(2)}(\boldsymbol{t}_0)]_1 + o(b^2) \\
& + O(n^{-\rho(\boldsymbol{t}_0)}) + O(n^{-2\delta(\boldsymbol{t}_0)}) + O_p(n^{-\delta(\boldsymbol{t}_0)} \sqrt{r_{nb}(\boldsymbol{t})}) + o\left((nb)^{-d/2}\right).
\end{aligned}
$$

□

Proof of Corollary 3.7

Proof. Let $\delta_{\boldsymbol{t}_1,\boldsymbol{t}_2} = E\left\{(W_n(\boldsymbol{t}_1) - \mu_n(\boldsymbol{t}_1))(W_n(\boldsymbol{t}_2) - \mu_n(\boldsymbol{t}_2))\right\}$. Then

$$
\begin{aligned}
& E\left\{ \sum_{\boldsymbol{t}_i \in [\boldsymbol{u},\boldsymbol{v}]^n} (W_n(\boldsymbol{t}_i) - \mu_n(\boldsymbol{t}_i))^2 - E(W_n(\boldsymbol{t}_i) - \mu_n(\boldsymbol{t}_i))^2 \right\}^x \\
= & \sum_{\boldsymbol{t}_{i_1},\ldots,\boldsymbol{t}_{i_x} \in [\boldsymbol{u},\boldsymbol{v}]^n} \sum_{\mathbf{S}} \delta_{\boldsymbol{t}_{s_1}\boldsymbol{t}_{s_2}} \delta_{\boldsymbol{t}_{s_3}\boldsymbol{t}_{s_4}} \cdots \delta_{\boldsymbol{t}_{s_{2x-1}}\boldsymbol{t}_{s_{2x}}},
\end{aligned}
$$

where $\mathbf{S}$ is the set of all possible ways to make x pairs, $\{(s_1, s_2), ..., (s_{2x-1}, s_{2x})\}$, with s_j chosen from $\mathbf{I}^x = \{i_1, i_1, i_2, i_2, ..., i_x, i_x\}$ without replacement and each pair must be of different indices. By assumption, $\delta_{\boldsymbol{t}_i,\boldsymbol{t}_j} \leq \mathbf{c}_5$ uniformly and from Lemma B.5, $\sum_{\boldsymbol{t}_i} \delta_{\boldsymbol{t},\boldsymbol{t}_i} \leq \mathbf{c}_{11} r_{nb}(\boldsymbol{t})$. Notice that if we let $\mathbf{c}_{23} = \max(1, \mathbf{c}_5, \mathbf{c}_{11})$, then as in the proof of Theorem 3.4, every chain of length q is bounded above by $\mathbf{c}_{23}^q (nb)^d s_{nb}(\boldsymbol{t}) r_{nb}(\boldsymbol{t})^{q-2}$. So therefore if x is even, $\left|\sum_{\boldsymbol{t}_{i_1},\ldots,\boldsymbol{t}_{i_x} \in [\boldsymbol{u},\boldsymbol{v}]^n} \delta_{\boldsymbol{t}_{s_1}\boldsymbol{t}_{s_2}} \delta_{\boldsymbol{t}_{s_3}\boldsymbol{t}_{s_4}} \cdots \delta_{\boldsymbol{t}_{s_{2x-1}}\boldsymbol{t}_{s_{2x}}}\right|$ is largest when each chain has length 2. Therefore

$$
\left| \sum_{\boldsymbol{t}_{i_1},\ldots,\boldsymbol{t}_{i_x} \in [\boldsymbol{u},\boldsymbol{v}]^n} \delta_{\boldsymbol{t}_{s_1}\boldsymbol{t}_{s_2}} \delta_{\boldsymbol{t}_{s_3}\boldsymbol{t}_{s_4}} \cdots \delta_{\boldsymbol{t}_{s_{2x-1}}\boldsymbol{t}_{s_{2x}}} \right| \leq \mathbf{c}_{23}^x (nb)^{xd/2} s(nb)^{x/2} \tag{B.36}
$$

Notice that the number of pairings in $\mathbf{S}$ is bounded above by $(2x-1)!!$. Combining with (B.36), for sufficiently large n

$$
\frac{(nb)^{-xd/2}}{s(nb)^{x/2}} E\left\{ \sum_{\boldsymbol{t}_i \in [\boldsymbol{u},\boldsymbol{v}]^n} (W_n(\boldsymbol{t}_i) - \mu_n(\boldsymbol{t}_i))^2 - E(W_n(\boldsymbol{t}_i) - \mu_n(\boldsymbol{t}_i))^2 \right\}^x \leq \mathbf{c}_{23}^x (2x-1)!!
$$

when x is even. For x odd we can apply Cauchy-Schwartz. Leting

$$G_n(\boldsymbol{u},\boldsymbol{v}) = \frac{(nb)^{-xd/2}}{s(nb)^{x/2}} \sum_{\boldsymbol{t}_i \in [\boldsymbol{u},\boldsymbol{v}]^n} (W_n(\boldsymbol{t}_i) - \mu_n(\boldsymbol{t}_i)) - E(W_n(\boldsymbol{t}_i) - \mu_n(\boldsymbol{t}_i)),$$

$$\begin{aligned}
E|G_n(\boldsymbol{u},\boldsymbol{v})|^x &= E\big\{|G_n(\boldsymbol{u},\boldsymbol{v})|^{\frac{x+1}{2}}|G_n(\boldsymbol{u},\boldsymbol{v})|^{\frac{x-1}{2}}\big\} \\
&\leq \sqrt{E(G_n(\boldsymbol{u},\boldsymbol{v}))^{x+1}E(G_n(\boldsymbol{u},\boldsymbol{v}))^{x-1}} \\
&\leq \sqrt{\mathbf{c}_{23}^{2x}(2x+1)!!(2x-1)!!} \\
&= \mathbf{c}_{23}^{x}(2x-1)!!\sqrt{2x+1} \\
&\leq 2^x\mathbf{c}_{23}^{x}(2x-1)!!,
\end{aligned}$$

where the last inequality holds since $\sqrt{2x+1} \leq 2^x$. So the theorem holds if we set $\mathbf{c}_7 = 2\mathbf{c}_{23}$. □

Proof of Theorem 3.8

Proof. Recall that $\hat{\beta}_0(\boldsymbol{t}_0; n, b) = \sum_{\boldsymbol{t}_i \in T_n(\boldsymbol{t}_0,b)} \tilde{k}_{\boldsymbol{t}_i} W_n^2(\boldsymbol{t}_i)$. Therefore the estimator we will consider is

$$\frac{1}{n^2b^2}\sum_{\boldsymbol{t}_i} \tilde{f}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right) W_n(\boldsymbol{t}_i)^2,$$

where $\tilde{f}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right) = n^2b^2\tilde{k}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right)$.

$$\begin{aligned}
&\frac{(nb)^{d/2}}{\log(n)V(n,b)}\left|\hat{\beta}_0(\boldsymbol{t}_0) - g(\boldsymbol{t}_0)\right| \\
=\ &\frac{(nb)^{d/2}}{\log(n)V(n,b)}\left|\sum_{\boldsymbol{t}_i\in[0,1]^d}\frac{1}{n^2b^2}\tilde{f}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right)W_n^2(\boldsymbol{t}_i) - g(\boldsymbol{t}_0)\right| \\
=\ &\frac{(nb)^{d/2}}{\log(n)V(n,b)}\left|\sum_{\boldsymbol{t}_i\in[0,1]^d}\frac{1}{n^2b^2}\tilde{f}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right)(\widetilde{W}_n^2(\boldsymbol{t}_i) + 2\mu_n(\boldsymbol{t}_i)\widetilde{W}_n(\boldsymbol{t}_i) + \mu_n^2) - g(\boldsymbol{t}_0)\right| \\
\leq\ &\frac{(nb)^{d/2}}{\log(n)V(n,b)}\left|\sum_{\boldsymbol{t}_i}\frac{1}{n^2b^2}\tilde{f}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right)\left\{\widetilde{W}_n^2(\boldsymbol{t}_i) - E\widetilde{W}_n^2(\boldsymbol{t}_i)\right\}\right| \qquad (B.37) \\
&+\frac{(nb)^{d/2}}{\log(n)V(n,b)}\left|\sum_{\boldsymbol{t}_i}\frac{1}{n^2b^2}\tilde{f}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right)2\mu_n(\boldsymbol{t}_i)\widetilde{W}_n(\boldsymbol{t}_i)\right|. \qquad (B.38) \\
&+\frac{(nb)^{d/2}}{\log(n)V(n,b)}\left|\sum_{\boldsymbol{t}_i}\frac{1}{n^2b^2}\tilde{f}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right)EW_n^2(\boldsymbol{t}_i) - g(\boldsymbol{t}_0)\right| \qquad (B.39)
\end{aligned}$$

From Theorem B.7 and Corollary B.8, since $\tilde{f}$ satisfies Assumption 3.6 and $\mu_n(\boldsymbol{t}) = O(n^{-\delta(\boldsymbol{t})})$,

$$\sup_{\boldsymbol{t}_0 \in [0,1]^2} \frac{(nb)^{d/2}}{\log(n) s_{nb}(\boldsymbol{t}_0)^{1/2}} \left| \sum_{\boldsymbol{t}_i} \frac{1}{n^2 b^2} \tilde{f}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right) \left\{ \widetilde{W}_n^2(\boldsymbol{t}_i) - E\widetilde{W}_n^2(\boldsymbol{t}_i) \right\} \right|,$$
$$\sup_{\boldsymbol{t}_0 \in [0,1]^2} \frac{(nb)^{d/2} n^{\delta(\boldsymbol{t}_0)}}{\log(n) r_{nb}(\boldsymbol{t}_0)^{1/2}} \left| \sum_{\boldsymbol{t}_i} \frac{1}{n^2 b^2} \tilde{f}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right) 2\mu_n(\boldsymbol{t}_i) \widetilde{W}_n(\boldsymbol{t}_i) \right|$$

are less than or equal to $\mathbf{c}_{12}$ eventually w.p.1. Then since $V_{n,b}$ is asymptotically greater than or equal to $s_{nb}(\boldsymbol{t})^{1/2}$ and $r_{nb}(\boldsymbol{t})^{1/2}/n^{\delta(\boldsymbol{t})}$ uniformly, $\sup_{\boldsymbol{t}_0 \in [0,1]^d}$ of (B.37) and (B.38) are less than or equal to $\mathbf{c}_{12}$ eventually w.p.1.

Then from Theorem 3.3

$$\begin{aligned} & \frac{(nb)^{d/2}}{\log(n) V(n,b)} \sup_{\boldsymbol{t}_0 \in [0,1]^2} |E\hat{\beta}_0(\boldsymbol{t}_0; n, b) - g(\boldsymbol{t}_0)| \\ = \; & O\left(\frac{n^{d/2} b^{d/2+2}}{\log(n) V(n,b)}\right) + O\left(\frac{n^{d/2-\rho(\boldsymbol{t}_0)} b^{d/2}}{\log(n) V(n,b)}\right) + O\left(\frac{n^{d/2-2\delta(\boldsymbol{t}_0)} b^{d/2}}{\log(n) V(n,b)}\right) \end{aligned}$$

uniformly. Therefore for (B.39), we can choose n, b such that

$$\frac{(nb)^{d/2}}{\log(n) V(n,b)} \sup_{\boldsymbol{t}_0 \in [0,1]^2} \left| \sum_{\boldsymbol{t}_0 i \in [0,1]^d} \frac{1}{n^2 b^2} \tilde{f}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}_0}{b}\right) W_n^2(\boldsymbol{t}_i) - g(\boldsymbol{t}) \right| \leq \mathbf{c}_{12}$$

eventually w.p.1. Letting $\mathbf{c}_8 = 3\mathbf{c}_{12}$ completes the proof.

□

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] E. B. Anderes. On the consistent separation of scale and variance for gaussian random fields. *Ann. Statist.*, 38:870–893, 2010.

[2] E. B. Anderes and M. L. Stein. Estimating deformations of isotropic gaussian random fields on the plane. *Ann. Statist.*, 36:719–741, 2008.

[3] A. Ayache, S. Cohen, and J. Lévy Véhel. The covariance structure of multifractional brownian motion, with application to long range dependence. *ICASSP*, 2000.

[4] G. Baxter. A stong limit theorem for gaussian processes. *Proc. Amer. Math. Soc.*, 7:522–527, 1956.

[5] J.-P. Chilés and P. Delfiner. *Geostatistics*. Wiley, New York, 1999.

[6] M. Clerc and S. Mallat. Estimating deformations of stationary processes. *Ann. Stat.*, 31:1772–1821, 2003.

[7] N. Cressie. *Statistics for Spatial Data*. Wiley, New York, 1993.

[8] J. Fan, T. Gasser, I. Gijbels, M. Brockmann, and J. Engel. Local polynomial regression: optimal kernels and asymptotic minimax efficiency. *Ann. Statist.*, 49:77–99, 1997.

[9] E. G. Gladyshev. A new limit theorem for stochastic processes with gaussian increments. *Theor. Probability Appl..*, 6:52–61, 1961.

[10] E. Herbin. From n parameter fractional brownian motions to n parameter multifractional brownian motions. *Rocky Mountain Journal of Mathematics*, 3:1249–1284, 2006.

[11] J. Istas and G. Lang. Quadratic variations and estimation of the local hlder index of a gaussian process. *Ann. Inst. H. Poincaré Probab. Statist.*, 33:407–436, 1997.

[12] F. Kozin. A limit theorem for processes with stationary inependent increments. *Proc. Amer. Math. Soc.*, 8:960–963, 1957.

[13] P. Lévy. Le mouvement brownien plan. *Amer. J. Math.*, 62:487–550, 1940.

[14] E. A. Nadaraya. On estimating regression. *Theory Probab. Appl.*, 9:141–142, 1964.

[15] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *Ann. Statist.*, 27:832–837, 1956.

[16] P. D. Sampson and P. Guttorp. Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87:108–119, 1992.

[17] M. Stein. *Interpolation of Spatial Data*. Springer, Heidelberg, New York, 1999.

[18] C. J. Stone. Consistent nonparametric regression (with discussion). *Ann. Statist.*, 5:595–645, 1977.

[19] C. J. Stone. Optimal global rates of convergence for nonparametric regression. *Ann. Statist.*, 10:1040–1053, 1982.

[20] G. S. Watson. Smooth regression analysis. *Sankhyā Ser. A*, 26:359–372, 1964.