ARTICLE IN PRESS

Journal of the Korean Statistical Society xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

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

Variance function estimation of a one-dimensional nonstationary process

Eunice J. Kim [a,b,*], Zhengyuan Zhu [b]

[a] 1 Microsoft Way, Redmond, WA 98052, USA
[b] Iowa State University, Snedecor Hall, Ames, IA 50011, USA

ARTICLE INFO

Article history:
Received 9 July 2017
Accepted 7 January 2019
Available online xxxx

AMS 2000 subject classifications:
primary 62G05
secondary 62G20
62M30

Keywords:
Difference-based
Nonstationary process
Correlated errors
Variance function estimation

ABSTRACT

We propose a flexible nonparametric estimation of a variance function from a one-dimensional process where the process errors are nonstationary and correlated. Due to nonstationarity a local variogram is defined, and its asymptotic properties are derived. We include a bandwidth selection method for smoothing taking into account the correlations in the errors. We compare the proposed difference-based nonparametric approach with Anderes and Stein(2011)'s local-likelihood approach. Our method has a smaller integrated MSE, easily fixes the boundary bias, and requires far less computing time than the likelihood-based method.

1. Introduction

The prevalence of mobile devices and increase in storage capacity have brought about high demand for spatial data analysis. Many spatial processes exhibit nonstationary features, such as non-constant mean, variance, and autocorrelation. We encounter these features in many domains, and most commonly from sociology, ecology, geology, meteorology, and astronomy. Even in one-dimensional processes nonstationarity is common. Consider estimating a range of travel times to move from point A to point B. The mean and the variance of speed on each section of a road are non-constant, and the nearby locations share similar features than further apart locations. Consider setting up a wind turbine facing an optimal direction: the wind direction at a fixed location over time exhibits non-constant variance and is autocorrelated in short time span. In such scenarios, it is useful to construct interval estimates of the trend and provide spatial prediction intervals using the variance function estimation.

We propose a difference-based variance function estimator for a one-dimensional process where the errors are nonstationary and correlated. Prior to our method, the differencing has been applied to data with independent errors. Neumann, Kent, Bellinson, and Hart (1941) proposed using differences of successive observations to estimate the variance of independent and identically distributed (*i.i.d.*) errors. Seifert, Gasser, and Wolf (1993) and Wang, Brown, Cai, and Levine (2008) explored the reduction of the bias in the estimation of variance when differencing from skipping the estimation of the mean function, which is a source of bias. Gasser, Sroka, and Jennen-Steinmetz (1986) proposed second-order differencing to estimate variance functions when observations are irregularly spaced, and Hall, Kay, and Titterington (1990, 1991) used a differencing approach in image processing to estimate the variance of two-dimensional processes with *i.i.d.* errors.

* Corresponding author.
E-mail address: eunice.kim@microsoft.com (E.J. Kim).

https://doi.org/10.1016/j.jkss.2019.01.001
1226-3192/

In estimating the variance function of a one-dimensional spatial process where the mean and the variance functions are smooth with additive correlated errors, Anderes and Stein (2011) proposed a likelihood-based method. This method can handle irregularly spaced data and provide statistical efficiency when the Gaussian assumption is tenable for the observed process. Still, when selecting a smoothing bandwidth using a local likelihood ratio test heuristic, the computational burden is heavy as the covariance matrix of every simulated process must be inverted. The assumption of the Gaussian errors is also stringent for many nonstationary processes. In signal processing, a band-pass filter provides a local variance estimation assuming that the trend changes slowly. The method is suitable for a second-order stationary error process but not for a nonstationary error process. Also, converting the output from the frequency domain back to time domain often introduces bias, whereas applying the difference-based variance estimation in the time domain introduces less bias.

We assume an equidistant design for a one-dimensional process and consider an infill asymptotic framework for a variance function estimator. Brown and Levine (2007) discussed the asymptotic properties of nonparametric variance estimators formed by differencing non-constant and independent errors. Cai and Wang (2008) extended an adaptive approach to a variance function estimation using wavelet transforms. This article covers a method applied to a nonstationary correlated error process and considers general cross-validation for bandwidth selection. In addition to the estimation of a variance function, the method estimates a short-range correlation structure in the data.

We define a local variogram in Section 2 and describe its the asymptotic properties of the local variogram estimator in Section 3. Hall and Carroll (1989) discussed the asymptotic risk of the difference-based variance function estimator in nonparametric regression with regard to the smoothness of variance and mean functions, and Wang et al. (2008) derived the asymptotic minimax risk rate.

The paper is organized as follows. Section 2 defines a data model and local variogram as a product of variance and variogram functions. Section 3 proposes the estimator of local variogram and describes its theoretical properties. Section 4 presents the algorithm for the variance function estimation. Section 5 evaluates the method through a simulation study and discusses the advantages of the difference-based variance function estimator compared to the likelihood-based estimator. Section 6 closes with possible extensions.

2. Data model and definitions

In this section, we define a Lipschitz condition and our nonstationary data model and introduce local variogram. A Lipschitz condition on the mean and variance functions of the nonstationary data helps to define the estimable variance functions.

Definition 1 (*Lipschitz Condition*)**.** Let $c_1, c_2 > 0$. Denote $q' \doteq q - \lfloor q \rfloor$ where $\lfloor q \rfloor$ is the largest integer less than q. We say that the function $f(x)$ is in class of $\Lambda_q(c_f)$ if for all $x, y \in (0, 1)$, $\left|f^{(\lfloor q \rfloor)}(x) - f^{(\lfloor q \rfloor)}(y)\right| \le c_1 |x - y|^{q'}$, $\left|f^{(k)}(x)\right| \le c_2$ for $k = 0, \ldots, \lfloor q \rfloor$, and $c_f = \max(c_1, c_2)$.

Definition 2. If a function $f(x)$ is in class $\Lambda_q(c_f)$ and there exists $\delta > 0$ such that $f(x) > \delta$ for all $x \in [0, 1]$, we say the function is in $\Lambda_q^+(c_f)$.

Data Model Consider a nonstationary continuous process model

$$Z(s) = \mu(s) + \sigma(s)X(s) \tag{1}$$

on $0 \le s \le 1$ without loss of generality. We assume a smooth mean function $\mu(s)$ and an additive, correlated noise as a product of a smooth standard deviation function $\sigma(s)$ and a second-order stationary process $\{X(s)\}$ where $E(X(s)) = 0$, $var(X(s)) = 1$, and $cov(X(s), X(s')) = \rho(\left|s - s'\right|; \theta)$ for all pairs of s and s' in the unit interval. Consider $\mu(s) \in \Lambda_q(c_f)$, $q \ge 0$, and $\sigma^2(s) \in \Lambda_\beta^+(c_f)$, $\beta \ge 2$. The correlation function follows:

$$\rho(\left|s - s'\right|; \theta) = \begin{cases} 1 & s = s' \\ 1 - \dfrac{\left|s - s'\right|^\alpha}{\theta} + O(\left|s - s'\right|^{\alpha+2}) & s \ne s' \end{cases} \tag{2}$$

where $\theta > 0$ and $0 < \alpha < 2$ for validity. This class of correlation function (2) encompasses linear, spherical, Matérn and exponential models (Stein, 1999). For an equally spaced design, we define the location with $s_i = (2i - 1)/(2n)$ indexed by $i = 1, \ldots, n$. As a shorthand we write $Z_i = Z(s_i)$, $\mu_i = \mu(s_i)$, $\sigma_i = \sigma(s_i)$, $\rho_h = \rho(h/n)$ and specify a parametric correlation function $\rho_{s;\theta} = \rho(s; \theta)$. Let $\sigma^{2(j)}(s) = d^j\sigma^2(x)/dx^j\big|_{x=s}$ denote the jth-order derivative of a function $\sigma^2(s)$.

We expand the definition of a *variogram*, introduced by Matheron (1962), since differencing a nonstationary process depends on its local properties. Using a 0-mean nonstationary process as a data model from (1), the variance at s of a lag-$\frac{h}{n}$ first-order differenced process is

$$\begin{aligned} &var\left(Z\left(s - \frac{h}{2n}\right) - Z\left(s + \frac{h}{2n}\right)\right) \\ &= 2\sigma^2(s)(1 - \rho_h) + 2\left(\sigma^{(1)}(s)\right)^2(1 + \rho_h)\left(\frac{h}{2n}\right)^2 + o\left(n^{-2}\right). \end{aligned} \tag{3}$$

Please cite this article as: E.J. Kim and Z. Zhu, Variance function estimation of a one-dimensional nonstationary process. Journal of the Korean Statistical Society (2019), https://doi.org/10.1016/j.jkss.2019.01.001.

The first term contains the product of a variogram and a local variance. The second and following higher order terms are comprised of the derivatives of the local variance function and the power of lag.

Definition 3. The local variogram $2\gamma_L(s, h; \theta)$ is defined as the leading term of (3), i.e.

$$\gamma_L(s, h; \theta) = \sigma^2(s)\left(1 - \rho\left(\frac{h}{n}; \theta\right)\right). \tag{4}$$

Local variogram in (4) is a product of a variance function and the variogram of a stationary process. While a variogram represents spatial dispersion by taking lagged differences of a stationary process, a local variogram describes the spatial dispersion about a specific neighborhood. When the lag size h is small in comparison to the number of observed points n in mixed-domain asymptotic, the higher order terms in (3) vanish. With an increasing domain, the variance function is better estimated. With an infill asymptotic, the correlation is better described.

3. Theoretical results

To estimate a variance function, we first define an estimator for local variogram in Section 3.1. The bias and the variance of the local variogram estimator are derived in Sections 3.2 and 3.3 respectively. In Section 3.4, the asymptotic rate of convergence of the point-wise mean square error is shown and is compared to that of the standard nonparametric variance estimator with i.i.d. errors.

3.1. Local variogram estimator

For an equally-spaced, nonstationary process $\{Z_i\}$ in (1), consider taking a simple differencing of lag $\frac{h}{n}$. Let $D_{i,h} = \frac{Z(s_i) - Z(s_{i+h})}{\sqrt{2}}$. We normalize the simple differenced process such that $var(D_{i,h})$ matches $var(Z_i)$ as if $\{Z_i\}$ were an independent process. We refer to the sequence $\{D_{i,h}\}_{i=1}^{n-h}$ as pseudo-residuals, borrowing the term from Brown and Levine (2007). The shape of squared pseudo-residuals at different lags resemble a variogram cloud.

Let K_λ represent a Gasser–Müller kernel with bandwidth λ

$$K_{\lambda,s_0}(s) = \sum_{i=0}^{k-2} a_i \left(\frac{s - s_0}{\lambda}\right)^i$$

$$\text{where } a_i = \begin{cases} 0 & (k+i) \text{ odd} \\ \frac{(-1)^{i/2}(k)!(k+i)!k(k-i)}{i!(i+1)2^{2k+1}+\left(\frac{k}{2}\right)!\left(\frac{k}{2}\right)!\left(\frac{k-i}{2}\right)!\left(\frac{k+i}{2}\right)!} & (k+i) \text{ even.} \end{cases}$$

Since we directly estimate the variance function, select the order of derivative $\nu = 0$ for the above Gasser–Müller kernel, and set the polynomial order k to be greater than the degree differentiability β of a variance function. Gasser, Müller, and Mammitzsch (1985) developed the kernel so that the moment conditions simplify the calculation of high-order terms in nonparametric estimators and that the edge effect be easily removed by adjusting the kernels at the boundaries of the domain.

Define the Gasser–Müller kernel estimator of local variogram at location s and lag $\frac{h}{n}$ as

$$\hat{\gamma}_{L\lambda}(s, h) = \sum_{i=1}^{n-h} K_{\lambda,i+h/2}(s) D_{i,h}^2. \tag{5}$$

Note that the ith squared difference $D_{i,h}^2$ is associated with the kernel weight centered at $s_{i+h/2}$ since the ith pseudo-residual is positioned directly between s_i and s_{i+h}. It is possible to consider higher-order differencing, but the first-order differencing introduces the least bias and variance in local variogram estimation due to the reduced number of correlated terms involved. We also suggest using the smallest lag in differencing because it reduces correlation among the sequence of pseudo-residuals $\{D_{i,h}^2\}$.

3.2. Bias in the local variogram estimator

Let $D_{i,h} = (Z_i - Z_{i+h})/\sqrt{2}$, $\delta_{i,h} = \mu_i - \mu_{i+h}$, and $g_{i,h} = \sigma_i^2 + \sigma_{i+h}^2 - 2\sigma_i\sigma_{i+h}\rho_h$ for $i = 1, \ldots, n-h$. The expected value of the local variogram estimator is

$$\begin{aligned} E\left(\hat{\gamma}_{L\lambda}(s, h)\right) &= \sum_{i=1}^{n-h} K_{\lambda,i+\frac{h}{2}}(s) E\left(D_{i,h}^2\right) \\ &= \frac{1}{2}\sum_{i=1}^{n-h} K_{\lambda,i+\frac{h}{2}}(s)\left\{(\mu_i - \mu_{i+h})^2 + \sigma_i^2 + \sigma_{i+h}^2 - 2\sigma_i\sigma_{i+h}\rho_h\right\}. \end{aligned}$$

Please cite this article as: E.J. Kim and Z. Zhu, Variance function estimation of a one-dimensional nonstationary process. Journal of the Korean Statistical Society (2019), https://doi.org/10.1016/j.jkss.2019.01.001.

The bias of the local variogram estimator is

$$\begin{aligned} bias(\hat{\gamma}_\lambda(s,h)) &= E(\hat{\gamma}_\lambda(s,h)) - (1-\rho_h)\sigma^2(s) \\ &= \sum_{i=1}^{n-h} K_{\lambda,i+\frac{h}{2}}(s)\left\{\frac{1}{2}(\delta_{i,h}^2 + g_{i,h}) - (1-\rho_h)\sigma^2(s)\right\}. \end{aligned} \tag{6}$$

Note that $(1-\rho_h) = O(n^{-\alpha})$ and $0 < \alpha < 2$.

Theorem 3.1. *Assume a nonstationary data model (1) and the correlation function (2). The mean and the variance functions $\mu(s)$ and $\sigma^2(s)$ are continuously differentiable Lipschitz functions (see Definitions 1 and 2) where $\mu(s) \in \Lambda_q(c_f)$, $q \geq 0$ and $\sigma^2(s) \in \Lambda_\beta^+(c_f)$, $\beta \geq 2$. Using the Gasser–Müller kernel, the local variogram estimator (5) at location s and lag $\frac{h}{n}$ has an asymptotic bias of order*

$$bias(\hat{\gamma}_\lambda(s,h)) = \begin{cases} O(n^{-2} + n^{-2q} + n^{-\alpha-1}) & where \quad q, \beta < m \\ O(n^{-2} + n^{-2q} + n^{-\alpha-1}) + O(n^{-\alpha}\lambda^m) & where \quad q < m \leq \beta \\ O(n^{-2} + n^{-2q} + n^{-\alpha-1}) + O(\lambda^m) & where \quad m \leq q. \end{cases} \tag{7}$$

where m is the order of kernel.

Proof. To calculate an asymptotic bias we split (6) into two parts. The first term is $\delta_{i,h}^2$ whose expansion is in (20) for $q \geq 1$ and in (21) for $0 \leq q < 1$. Convolved with a Gasser–Müller kernel of order m (Gasser et al., 1985), the higher order terms in $\delta_{i,h}^2$ cancel when the number of derivatives of the mean function $q \leq m$ and shows

$$\sum_{i=1}^{n-h} K_{\lambda,i+\frac{h}{2}}(s)\delta_{i,h}^2 = \begin{cases} O(n^{-2}) + O\left(n^{-2q}\right) & \text{where} \quad q < m \\ O(n^{-2}) + O\left(n^{-2q}\right) + O(\lambda^m) & \text{where} \quad q \geq m. \end{cases} \tag{8}$$

The second part of the bias is $\frac{1}{2}g_{i,h} - \sigma^2(s)(1-\rho_h)$. The leading term in the expansion of $g_{i,h}$ about s is the local variogram $\sigma^2(s)(1-\rho_h)$. See Eq. (22) in Appendix for the Taylor expansion. Applying the Gasser–Müller kernel to the high order terms of $g_{i,h}$, we have the following:

$$\begin{aligned} &\sum_{i=1}^{n-h} K_{\lambda,i+\frac{h}{2}}(s)\left\{\frac{1}{2}g_{i,h} - \sigma^2(s)(1-\rho_h)\right\} \\ &= \sum_{i=1}^{n-h} K_{\lambda,i+\frac{h}{2}}(s)\left\{(1-\rho_h)\left(\sigma_s\sigma_s^{(1)}\frac{h}{n} + \frac{\sigma_s\sigma_s^{(2)}}{2}\frac{h^2}{n^2}\right) + \frac{1}{2}\left(\sigma_s^{(1)}\frac{h}{n}\right)^2\right\} \\ &\quad + \sum_{i=1}^{n-h} K_{\lambda,i+\frac{h}{2}}(s)(1-\rho_h)\sum_{j=1}^{\lfloor\beta\rfloor}\left\{\frac{\left(\sigma_s^2\right)^{(j)}}{j!} + \frac{\left(\sigma_s^2\right)^{(j+1)}}{2(j+1)!}\left(1+\frac{h}{n}\right)\frac{h}{n}\right\}(s_i - s)^j \\ &\quad + \sum_{i=1}^{n-h} K_{\lambda,i+\frac{h}{2}}(s)\frac{h^2}{2n^2}\sum_{k=1}^{\lfloor\beta\rfloor}\sum_{j=1}^{k+1} c_k\sigma_s^{(j)}\sigma_s^{(k-j+2)}(s_i - s)^k + \sum_{i=1}^{n-h} K_{\lambda,i+\frac{h}{2}}(s)O(|s_i - s|^\beta) \\ &= \begin{cases} O\left(n^{-\alpha-1}\right) + O\left(n^{-2}\right) & \text{where } \beta < m \\ O\left(n^{-\alpha-1}\right) + O\left(n^{-2}\right) + O(n^{-\alpha}\lambda^m) & \text{where } \beta \geq m. \end{cases} \end{aligned} \tag{9}$$

Combine the results in (8) and (9), and we have the asymptotic bias of the local variogram. □

The order of bias is dependent on the differentiability of the mean and the variance functions q and β respectively, the order m of the kernel, and the smoothness α of the nonstationary process. When m is greater than both q and β, which is Case A in Remark 4 of the Appendix, the asymptotic bias is the smallest. Therefore, we recommend choosing a high order kernel function since q and β are unknown. See Remark 4 for other conditions.

3.3. Variance of the local variogram estimator

The variance of the local variogram estimator at location s and lag $\frac{h}{n}$ is

$$var(\hat{\gamma}_\lambda(s,h)) = \sum_{i=1}^{n-h}\sum_{j=1}^{n-h} K_{\lambda,i+\frac{h}{2}}(s)K_{\lambda,j+\frac{h}{2}}(s)cov(D_{i,h}^2, D_{j,h}^2). \tag{10}$$

Recall that $D_{i,h} = (\delta_i + \sigma_i X_i - \sigma_{i+h}X_{i+h})/\sqrt{2}$ where X_i is a stationary process with mean 0, variance 1, and a correlation function $cov(X_i, X_{i+h}) = \rho_h$. Let $\{X_i\}_{i=1}^n$ be a Gaussian process. Then $(\sigma_i X_i - \sigma_{i+h}X_{i+h})$ is distributed $Normal\left(0, g_{i,h}\right)$, and its

fourth moment is $E\,(\sigma_i X_i - \sigma_{i+h}X_{i+h})^4 = 3g_{i,h}^2$. The variance of the squared pseudo-residual is

$$\begin{aligned} var(D_{i,h}^2) &= E(D_{i,h}^4) - E^2(D_{i,h}^2) \\ &= \frac{1}{4}\left\{\delta_{i,h}^4 + 6\delta_{i,h}^2 g_{i,h} + 3g_{i,h}^2 - \left(\delta_{i,h}^2 + g_{i,h}\right)^2\right\} \\ &= \delta_{i,h}^2 g_{i,h} + \frac{1}{2}g_{i,h}^2. \end{aligned}$$

The covariance between the ith and the jth squared differences is

$$\begin{aligned} &cov(D_{i,h}^2, D_{j,h}^2) \\ &= \frac{1}{4}\left\{E((Z_i - Z_{i+h})^2\left(Z_j - Z_{j+h}\right)^2) - (\delta_{i,h}^2 + g_{i,h})(\delta_{j,h}^2 + g_{j,h})\right\} \\ &= \delta_{i,h}\delta_{j,h}\{\rho_{|i-j|}(\sigma_i\sigma_j + \sigma_{i+h}\sigma_{j+h}) - \rho_{|i-j-h|}\sigma_i\sigma_{j+h} - \rho_{|i-j+h|}\sigma_{i+h}\sigma_j\} \\ &\quad + \frac{1}{2}\{(\rho_{|i-j|}\sigma_i\sigma_j - \rho_{|i-j-h|}\sigma_i\sigma_{j+h})^2 + \left(\rho_{|i-j+h|}\sigma_{i+h}\sigma_j - \rho_{|i-j|}\sigma_{i+h}\sigma_{j+h}\right)^2\} \\ &\quad + (\rho_{|i-j|}^2 + \rho_{|i-j-h|}\rho_{|i-j+h|})\sigma_i\sigma_{i+h}\sigma_j\sigma_{j+h} - \rho_{|i-j|}\sigma_i\sigma_{i+h}(\rho_{|i-j+h|}\sigma_j^2 + \rho_{|i-j-h|}\sigma_{j+h}^2) \\ &= \delta_{i,h}\delta_{j,h}P_{ij} + \frac{1}{2}P_{ij}^2. \end{aligned}$$

where $P_{ij} = \rho_{|i-j|}\left(\sigma_i\sigma_j + \sigma_{i+h}\sigma_{j+h}\right) - \rho_{|i-j-h|}\sigma_i\sigma_{j+h} - \rho_{|i-j+h|}\sigma_{i+h}\sigma_j$ for $i \neq j$ and $P_{ii} = g_{i,h}$ for $i = j$. The Taylor expansion of $P_{i,j}$ about s_i for any $i \neq j$ is

$$P_{ij} = \frac{h^2}{n^2}\left(\sigma_i^{(1)}\right)^2 - \frac{2h^2}{(n\theta)^2}\sigma_i^2 + o\left(n^{-3}\right). \tag{11}$$

The next theorem shows the asymptotic rate of convergence of the variance of local variogram estimator.

Theorem 3.2. *Assume the same conditions as in Theorem 3.1 and a Gaussian process for $\{Z_i\}$. The variance (10) of the local variogram estimator $\hat{\gamma}_{L,\lambda}(s,h)$ in (5) is asymptotically*

$$var(\hat{\gamma}_\lambda(s,h)) = O\left(\frac{1}{n\lambda}\right)O\left(n^{-2q-\alpha} + n^{-2\alpha}\right). \tag{12}$$

Proof. Use the Taylor expansions of $\delta_{i,h}$, $g_{i,h}$, and P_{ij} in (11) (further details are in Eqs. (20) and (22) in the Appendix), and obtain the Taylor expansion of the variance about s at fixed lag $\frac{h}{n}$:

$$\begin{aligned} var(\hat{\gamma}_\lambda(s,h)) &= \sum_{i=1}^{n-h} K_{\lambda,i+\frac{h}{2}}^2(s)\left(\delta_i^2 g_i + \frac{g_i^2}{2}\right) + 2\sum_{i>j=1}^{n-h-1} K_{\lambda,i+\frac{h}{2}}(s)K_{\lambda,j+\frac{h}{2}}(s)\left(\delta_i\delta_j P_{ij} + \frac{P_{ij}^2}{2}\right) \\ &= 2\sum_{i=1}^{n-h} K_{\lambda,i+\frac{h}{2}}^2(s)\left\{\delta_i^2(1-\rho_h)O(1) + (1-\rho_h)^2 O(1)\right\} \\ &\quad + 2\sum_{i>j=1}^{n-h-1} K_{\lambda,i+\frac{h}{2}}(s)K_{\lambda,j+\frac{h}{2}}(s)\left\{\delta_i\delta_j O(n^{-2}) + O(n^{-4})\right\} \\ &= 2(1-\rho_h)\sum_{i=1}^{n-h} K_{\lambda,i+\frac{h}{2}}^2(s)\left\{O(n^{-2}+n^{-2q}) + (1-\rho_h)O(1)\right\} \\ &\quad + 2\sum_{i>j=1}^{n-h-1} K_{\lambda,i+\frac{h}{2}}(s)K_{\lambda,j+\frac{h}{2}}(s)O(n^{-4}) \end{aligned} \tag{13}$$

Use the fact that $K_{\lambda,i+\frac{h}{2}} = O\left(\frac{1}{n\lambda}\right)$ and $\sum K_{\lambda,i+\frac{h}{2}}^2 = O\left(\frac{1}{n\lambda}\right)$ and reduce the last line (13) to (12). □

The correlation between $D_{i,h}^2$ and $D_{j,h}^2$, where $i \neq j$, is

$$cor(D_{i,h}^2, D_{j,h}^2) = \frac{cov(D_{i,h}^2, D_{j,h}^2)}{\sqrt{var(D_{i,h}^2)var(D_{j,h}^2)}}$$

$$= \frac{\delta_{i,h}\delta_{j,h}P_{ij} + \frac{1}{2}P_{ij}^2}{\sqrt{(\delta_{i,h}^2 g_{i,h} + \frac{1}{2}g_{i,h}^2)(\delta_{j,h}^2 g_{j,h} + \frac{1}{2}g_{j,h}^2)}}$$

$$= \frac{\frac{h^4}{n^4}\left[\frac{2\sigma_i^2}{\theta^2}\left\{\frac{\sigma_i^2}{\theta^2} - \left(\sigma_i^{(1)}\right)^2 - \delta_{i,h}\delta_{j,h}\frac{n^2}{h^2}\right\} + \left\{\delta_{i,h}\delta_{j,h}\frac{n^2}{h^2} + \frac{1}{2}\left(\sigma_i^{(1)}\right)^2\right\}\left(\sigma_i^{(1)}\right)^2 + o\left(n^{-1}\right)\right]}{\sqrt{\left(\delta_{i,h}^2 g_{i,h} + \frac{1}{2}g_{i,h}^2\right)\left(\delta_{j,h}^2 g_{j,h} + \frac{1}{2}g_{j,h}^2\right)}}$$

$$= \frac{O(n^{-4})}{O(n^{-2\alpha})} = O(n^{-2(2-\alpha)}).$$

The correlation model of $\{Z_i\}$ in (2) sets $0 < \alpha < 2$. Note that the correlation between the squared pseudo-residuals $D_{i,h}^2$ and $D_{j,h}^2$ converges to 0 as $n \to \infty$ and the distance $\frac{|i-j|}{n}$ shortens. The speed of convergence is slower where α approaches 2, which translates to a very smooth process, which is a rarity in any physical process, assuming that $\{Z_i\}$ is a Gaussian process (Stein, 1999).

3.4. Asymptotic risk

Let the point-wise risk of the local variogram estimator be the point-wise sum of the squared bias in (6) and the variance in (13). The asymptotic point-wise risk combines the results of Theorems 3.1 and 3.2. We use $\asymp$ to represent the order of bandwidth, λ, in n.

Theorem 3.3. *Consider estimating the variance function of a one-dimensional nonstationary process with n equally-spaced observations, whose data model follows (1), (2), and a Gaussian distribution. Assume that $\mu(s) \in \Lambda_q$, $q \geq 0$, $\sigma^2(s) \in \Lambda_\beta$, $\beta \geq 2$ and that the bandwidth $\lambda = O\left(n^{-x}\right)$ where $0 < x < 1$. When the order of Gasser–Müller kernel m is $1 < m < \beta$ regardless of q, or $\beta < m < q$, the point-wise risk of the estimator of local variogram in (5) is*

$$Risk(\hat{\gamma}_\lambda(s,h)) = \begin{cases} O\left(n^{-4q}\right) & where \quad \lambda \asymp n^{-1-2\alpha+4q} \\ O\left(n^{-4}\right) & where \quad \lambda \asymp n^{3-2\alpha} \end{cases} \tag{14}$$

given $\alpha < 2q < \min(\alpha + \frac{1}{2}, 2)$ for the top case and $q \geq 1$ and $\alpha > \frac{3}{2}$ for the bottom case. When the order of Gasser–Müller kernel m is greater than either $q > 1$ or β, the point-wise risk is

$$Risk(\hat{\gamma}_\lambda(s,h)) = \begin{cases} O\left(n^{-2m(1+2\alpha)/(1+2m)}\right) & where \quad \lambda \asymp n^{-(1+2\alpha)/(1+2m)} \\ O\left(n^{-2\alpha-2m/(1+2m)}\right) & where \quad \lambda \asymp n^{-1/(1+2m)} \end{cases} \tag{15}$$

given $\alpha < \min\left(2q, \frac{3}{2}\right)$ for the top and $\alpha < 2q$ for the bottom.

Proof. The asymptotic bias and variance are derived in (7) and (12) respectively. Combining the two yields

$$Risk\left(\hat{\gamma}_\lambda(s,h), \gamma(s,h)\right) = bias(\hat{\gamma}_\lambda(s,h))^2 + var\left(\hat{\gamma}_\lambda(s,h)\right)$$

$$= \begin{cases} O\left(n^{-4} + n^{-4q}\right) + O\left(\frac{1}{n\lambda}\right) O\left(n^{-2\alpha}\right) & \text{where} \quad q, \beta < m \\ O\left(n^{-4} + n^{-4q} + n^{-2\alpha}\lambda^{2m}\right) + O\left(\frac{1}{n\lambda}\right) O\left(n^{-2\alpha} + n^{-2q-\alpha}\right) & \text{where} \quad q < m \leq \beta, \\ O(n^{-4} + n^{-4q} + \lambda^{2m}) + O\left(\frac{1}{n\lambda}\right) O\left(n^{-2\alpha} + n^{-2q-\alpha}\right) & \text{where} \quad m \leq q. \end{cases} \tag{16}$$

We break down the above three scenarios as A, B, and C below.

A. Assume that $m > q$ and $m > \beta$.

(i) When $q \geq 1$, $\alpha < 2q$ holds true because $0 < \alpha < 2$, and (12) reduces to $O\left(n^{-2\alpha-1}\lambda^{-1}\right)$. When the asymptotic bias is $O(n^{-2})$, the bandwidth condition is met and $\lambda \asymp n^{3-2\alpha}$, which suggests $\frac{3}{2} < \alpha$.

(ii) When $\frac{1}{2} < q < 1$, the asymptotic order of bias is $O\left(n^{-2q}\right)$. With $\alpha < 2q$, the asymptotic variance of $O\left(n^{-2\alpha-1}\lambda^{-1}\right)$. Then, $\lambda \asymp n^{-1-2\alpha+4q}$.

Please cite this article as: E.J. Kim and Z. Zhu, Variance function estimation of a one-dimensional nonstationary process. Journal of the Korean Statistical Society (2019), https://doi.org/10.1016/j.jkss.2019.01.001.

B. Assume $q < m \leq \beta$.

- When $\alpha < 2q$, the asymptotic variance is $O\left(n^{-2\alpha-1}\lambda^{-1}\right)$.

(i) The bias is $O(n^{-\alpha}\lambda^m)$ when $\alpha < 2q$, and it gives $\lambda \asymp n^{-1/(1+2m)}$.

(ii) The bias is $O(n^{-2})$ when $q > 1, \alpha > 2 - \dfrac{m}{1+2m}$, and $\lambda \asymp n^{3-2\alpha}$.

(iii) The bias is $O\left(n^{-2q}\right)$ when $q \leq 1, \alpha > \dfrac{1}{2q - \frac{m}{1+2m}}$, and $\lambda \asymp n^{-1-2\alpha+4q}$.

- When $\alpha \geq 2q$, the case does not hold. The asymptotic variance is $O\left(n^{-2q-\alpha-1}\lambda^{-1}\right)$, yet the conditions of the bias contradict the assumption.

C. Assume $m \leq q$.

(i) Assuming that the bias is $O(\lambda^m)$, we have $\lambda \asymp n^{-(1+2\alpha)/(1+2m)}$ when $\alpha < 2q$; and $\lambda \asymp n^{-(1+\alpha+2q)/(1+2m)}$ when $\alpha \geq 2q$. Detailing further conditions,

- $O(\lambda^m) > O(n^{-2}) \Longleftrightarrow \dfrac{m(1+2\alpha)}{1+2m} < 2$. This implies $m < \dfrac{2}{2\alpha - 3}$ when $\alpha > \dfrac{3}{2}$.
- $O(\lambda^m) > O(n^{-2q}) \Longleftrightarrow \dfrac{m(1+2\alpha)}{1+2m} < 2q$. This holds true since $\dfrac{1+2\alpha}{1+2m} < \dfrac{1+4q}{1+2m} < \dfrac{2q}{m}$ where the second inequality holds when $m < 2q$.

(ii) When the bias is $O(n^{-2})$, the case is the same as A(i) and B(ii), and $\lambda \asymp n^{3-2\alpha}$. □

Remark 1. In (14) of Theorem 3.3, the order of risk and bandwidth are the same for the top and bottom cases when the mean function is once differentiable ($q = 1$).

Given $m = \beta$ and as $\alpha \to 0$ (which suggests an independent process), the risk converges to $O\left(n^{-2\beta/(1+2\beta)}\right)$ in both cases of (15). The rate of convergence of the risk is consistent with the nonparametric estimation of a continuous, β-differentiable function (Tsybakov, 2009).

Remark 2. In (15) of Theorem 3.3 where the order of Gasser–Müller kernel is greater than the degree differentiability of the mean or the variance function, as long as $\alpha < 3/2$ with $m > 1$ or as long as $m > 3/2$ the risks are in a similar order of magnitude:

$$\frac{O\left(n^{-2m(1+2\alpha)/(1+2m)}\right)}{O\left(n^{-2\alpha-2m/(1+2m)}\right)} = O\left(n^{2\alpha/(1+2m)}\right).$$

Remark 3. There is a divergence of risk when (i) $q \geq \beta$ or (ii) the process is very smooth with $\alpha \gtrsim 3/2$, as it is hard to distinguish the mean function from a nonstationary noise process.

4. Bandwidth selection

It is well known in the nonparametric statistics literature that with correlation in underlying data, a cross-validation for bandwidth selection requires an adjustment to the data or a penalty term included in an objective function. Opsomer, Wang, and Yang (2001) compiled several proposals of bandwidth selection in nonparametric regression with correlated errors and addressed recent developments on the theoretical front. We choose a generalized cross-validation to minimize the mean square prediction errors of local variogram.

Recall that $D^2_{i,h}$ denotes the ith squared difference of lag-$\frac{h}{n}$ process. Let $d^2_{i,h}$ represent a realization of $D^2_{i,h}$, and define a deviance of local variogram estimation at $s_{i+h/2}$ as

$$\hat{\epsilon}_i = d^2_{i,h} - \hat{\gamma}_L\left(s_{i+\frac{h}{2}}, h\right). \tag{17}$$

Let the covariance matrix of the deviances be C_ϵ whose (i, j) element is $cov\left(\epsilon_i, \epsilon_j\right)$. We de-correlate the sequence of deviances, $resid_{\hat{\epsilon}} = (\hat{\epsilon}_1, \ldots, \hat{\epsilon}_n)$, of the local variogram estimation and denote the de-correlated residuals as

$$\xi = C_{\hat{\epsilon}}^{-1/2} resid_{\hat{\epsilon}}. \tag{18}$$

The choice of a covariance model and parameter values are not sensitive to the bandwidth estimation when the correlation in $resid_{\hat{\epsilon}}$ is weak with a small lag in differencing.

Please cite this article as: E.J. Kim and Z. Zhu, Variance function estimation of a one-dimensional nonstationary process. Journal of the Korean Statistical Society (2019), https://doi.org/10.1016/j.jkss.2019.01.001.

Difference-based variance function estimation including a bandwidth selection:

1. Take a simple difference of lag $\frac{h}{n}$ from $\{Z_i\}$. Create a set of bandwidths $\{\lambda_k\}$ whose values do not exceed 1/2 the range of the sample domain $[s_1, s_n]$. Estimate the local variogram using Eq. (5) for each λ_k.[1]
2. Calculate the $resid_{\hat{\epsilon}}$ in (17) for each λ_k and derive a sample covariance $C_{\hat{\epsilon}}$ of $resid_{\hat{\epsilon}}$.
3. Select a bandwidth using a generalized cross-validation, minimizing the overall mean squared error
$$\hat{\lambda}^* \leftarrow \arg_\lambda \min \sum_{i=1}^{n-1} \left(\frac{\xi_i}{1 - M_{(i,i)}} \right)^2,$$
where M is an $(n-h) \times (n-h)$ smoothing matrix of $D^2_{i,h}$ and the $(i,\ i)$ element of M is $M_{(i,i)} = K_\lambda(0)$ and ξ in (17).
4. Estimate the trend $\mu(s)$ of $\{Z_i\}$ and normalize the nonstationary process with $\hat{\mu}(s)$ and the local variogram using $\hat{\lambda}^*$ from 3:
$$\left\{Z_i^*\right\}_{i=1}^n \leftarrow \frac{\left\{Z_i - \hat{\mu}_i\right\}_{i=1}^n}{\left\{\sqrt{\widehat{\gamma}_{L\,\hat{\lambda}^*}(s_i, h)}\right\}_{i=1}^n}.$$
5. Fit a covariance model for $\left\{Z_i^*\right\}$ and estimate its variance $\hat{\sigma}_*^2$ and the correlation function at lag $\frac{h}{n}$. Adjust the local variogram estimation by the ratio between the first two, $\hat{\sigma}_*^2$ and $1 - \hat{\rho}_{Z^*}(h; \hat{\theta})$, and derive the variance function of $\{Z_i\}$:
$$\hat{\sigma}^2(s) \leftarrow \frac{\widehat{\gamma}_{L\,\hat{\lambda}^*}(s; h)\,\hat{\sigma}_*^2}{1 - \hat{\rho}_{Z^*}(h; \hat{\theta})}.$$

The result of our simulation study is in the next section. The generalized cross-validation is an approximation to the leave-one-out cross-validation. In Step 3, we use a de-correlated series to measure the risk, mean squared error. In Step 4, the estimation of the trend is needed when $\{Z_i\}$ is comprised of both the nonstationary error process and the trend. In Step 5, the parameters may be estimated either nonparametrically or parametrically. The bandwidth selection takes a few steps for optimization (from Step 1 to 3), whereas in the likelihood-based method (Anderes & Stein, 2011) it requires compute-intensive simulations to rank the given nonstationary process amongst conditioned, stationary processes.

5. Simulation study

We compare the difference-based method and the likelihood-based method in terms of statistical and computational efficiencies. We also examine the size of dependence in correlated errors on the functional estimations. To start, define *oracle bandwidth* as the bandwidth that yields the minimum discretely integrated mean square error (DMSE), which is the sum of the MSEs at a set of evaluation points. To provide equal footing on the difference- and likelihood-based estimations, we assume that the correlation functions and the parameter values are known. We label the oracle bandwidths for each method as 'Diff-λ^O' and 'Like-λ^O'.

5.1. Set-up

Assume a data model $Z(s) = \mu(s) + \sigma(s)X(s)$ as in (1) and set $\mu(s) = 0$ to test the method directly on a correlated error process $\{\sigma(s)X(s)\}$. We set the stationary error process $\{X(s)\}$ as a Gaussian process for analytical tractability. A Gaussian process is easy to simulate and fits the assumed data model for the likelihood-based approach, while it provides little favor towards the difference-based approach. The dependent structure is generated using an exponential correlation function with a range parameter set at two levels $\theta = 0.01$ and 0.1. The latter, in fact, refers to an independent error process. The observations are taken from an equally spaced grid over a unit interval, $s \in [0, 1]$. Four sample sizes $n = 100, 200, 500$, and 1000 are used. The standard deviation functions are chosen to examine the effect of differentiability of the variance functions especially for the bandwidth selections. Here is the summary of experimental set-up. Note that Anderes and Stein (2011) used $\sigma(s)$ in 2(a), and we add 2(b).

1. $n = 100, 200, 500$ and 1000
2. $\sigma(s) : [0, 1] \to \mathbf{R}^+$ and set $s \in \left\{0, \frac{1}{n-1}, \ldots, \frac{n-2}{n-1}, 1\right\}$.
   (a) an infinitely-differentiable function: $\sigma(s) = 2\sin(s/0.15) + 2.8$,
   (b) a step function: $\sigma(s) = 1 + \mathbb{1}_{\{1/3 < s \le 1\}}$.

[1] An important consideration in both local variogram and a variance function estimation is that they are non-negative everywhere. When a bandwidth is small, the smoothing may result in negative values often near the boundaries. When estimating near the boundary, we suggest fixing the bandwidth.

Please cite this article as: E.J. Kim and Z. Zhu, Variance function estimation of a one-dimensional nonstationary process. Journal of the Korean Statistical Society (2019), https://doi.org/10.1016/j.jkss.2019.01.001.

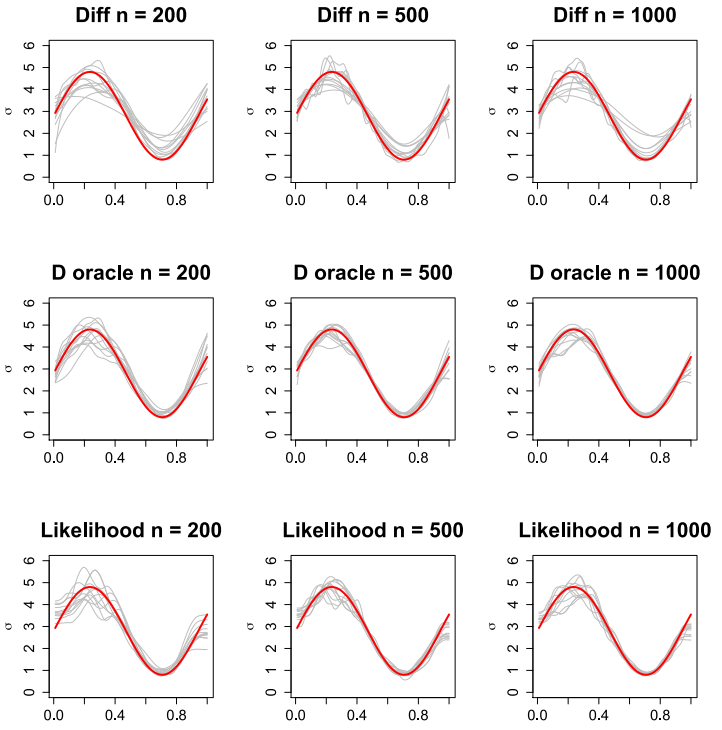

Fig. 1. A comparison of the difference-based method (first row) and the likelihood-based method (third row) using their respective proposed bandwidth selections. The middle row has the difference-based estimations using the bandwidth that minimizes the DMSE. The true standard deviation function is in thick red line, and the sample estimations are in thin gray lines. As the sample size increases from $n = 200$ to $n = 1000$, the overall estimation becomes more precise.

3. For a stationary error process, $\{X_s\}$, the correlation function is

$$cor\left(X(s), X\left(s+\frac{h}{n}\right)\right) = \begin{cases} 0 & \text{where } \theta = 0 \\ \exp\left(-\frac{1}{\theta}\frac{h}{n}\right) & \text{where } \theta > 0, \end{cases}$$

   and set $\theta = 0.01, 0.1$ for $h \ll n$ and $0 \le s \le 1 - \frac{h}{n}$.
4. Draw 100 random processes for each experimental setting.

Define $\text{DMSE}(\hat{\sigma}^2_{\hat{\lambda}}) = \sum_{i=1}^{n}(\hat{\sigma}_{i,\hat{\lambda}} - \sigma_i)^2/n$ and $L_\infty(\hat{\sigma}^2_{\hat{\lambda}}) = \max_i \left\{|\hat{\sigma}^2_{i,\hat{\lambda}} - \sigma_i^2|\right\}$. We estimate the variance functions at 100 equally spaced locations on [0,1] and evaluate using discretely integrated mean square error (DMSE) as an overall measure of functional estimation and the supremum norm L_∞, i.e. the maximum absolute deviation (MAX), to measure the worst discrepancy.

5.2. Experiments

Fig. 1 shows a series of results from the variance function estimation. The thick red line represents the true standard deviation function and the thin gray lines the estimations. The first row shows the proposed difference-based method (from here on noted *Diff*) including bandwidth and covariance parameter estimation; the second row applied the difference-based method with oracle bandwidths and known covariance parameters (noted *Diff*-λ^O); and the last row used the likelihood-based method (Anderes & Stein, 2011) (noted *Like*-λ^O) with oracle bandwidths and known covariance parameters. As expected in an infill design, the estimation becomes more precise as the number of observations increases from 200 to 500 to 1000. Comparing the second and third rows of Fig. 1, we see that the likelihood-based variance function estimations are more wavy than the difference-based ones. It suggests that the oracle bandwidths, that minimized the DMSE, for the likelihood-based method are underestimated. Had we selected a larger bandwidth for *Like*-λ^O to correct the shape of the functional estimations, the DMSE should only increase.

5.3. Numerical results

Figs. 2 and 3 display the estimation results of the sinusoidal $\sigma(\cdot)$ function with the discretely integrated mean square error (DMSE) and the maximum absolute deviation (MAX) respectively. The colors of the boxplots represent the estimation

Please cite this article as: E.J. Kim and Z. Zhu, Variance function estimation of a one-dimensional nonstationary process. Journal of the Korean Statistical Society (2019), https://doi.org/10.1016/j.jkss.2019.01.001.

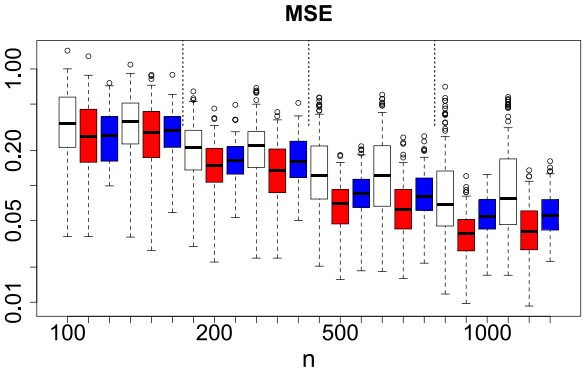

Fig. 2. Summary of the difference-based method (white) with correlation and bandwidth estimations, the same method with oracle correlation and bandwidths (red), and the likelihood-based method with oracle correlation bandwidths (blue) using DMSE. The sample size n is varied from 100 to 1000 on a fixed unit interval, and the strength of the error correlation is set weak (left) and strong (right) for each sample size. The y-axis is displayed in log-scale . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

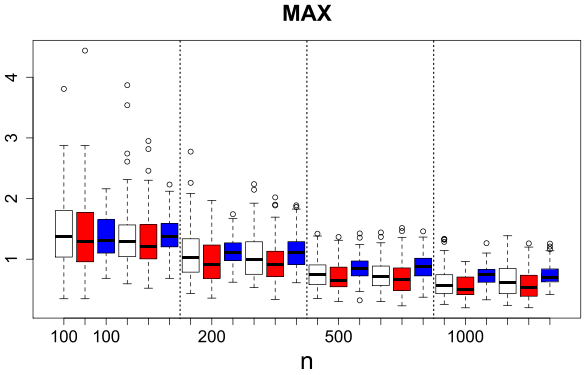

Fig. 3. Summary of three methods using the L_∞ norm, maximum absolute deviation. The detailed descriptions are the same as in Fig. 2.

methods where *Diff* is in white, the *Diff*-λ^O in red, and the *Like*-λ^O in blue. There are two sets of tri-colored boxplots for each sample size $n = 100, 200, 500$, and 1000 (demarcated by vertical dashed lines), where the left set corresponds to weakly correlated errors with $\theta = 0.01$ and the opposite set to strongly correlated errors with $\theta = 0.1$.

With oracle bandwidths, there is little difference between the difference-based and likelihood-based methods where n is less than 200. However, with larger sample sizes the difference-based method shows more consistency overall demonstrated by smaller DMSE and the supremum norm. The likelihood-based method with oracle bandwidth resulted in under-smoothed estimations of the variance function and exhibited the boundary effect (Fig. 1). The proposed difference-based method with a bandwidth estimation fares comparably to the difference-based estimation with an oracle bandwidth in the L_∞ norm (Fig. 3), but the risk of the estimation does not converge at the same rate as the applied method with an oracle bandwidth due to the variability resulting from the correlation estimation. In other words, the rate of convergence in the variance function estimator follows that of the local variogram estimator in Section 3 when the correlation structure is accurately estimated.

Considering that the sinusoidal $\sigma(\cdot)$ ranges from 0.8 to 4.8 in Fig. 1, the spread of summary values by both DMSE and the supremum norm is reasonable. In Fig. 2 the DMSEs are mostly less than 0.5, and in Fig. 3 the L_∞ norms are generally less than 1.5. Also note that the strength of the dependency in the error process or the effective range of correlation does not affect the asymptotic risk as displayed by similar distributions of the summary measures in the left and right triplets for each n.

5.4. Bandwidth selection

Table 1 contains the summary of bandwidth selections for estimating a smooth sinusoidal and piece-wise linear $\sigma(\cdot)$ functions. For smoothing kernels we use a degree six Gasser–Müller kernel for the differenced-based method and a Gaussian-based higher order kernel for the likelihood-based method. When we estimate both the variance and correlation functions, the bandwidth selection results in a large value in comparison to the oracle bandwidths with known correlation parameters. When there are more unknowns in the form of a process, a larger bandwidth smooths the relative instability in the estimation.

Please cite this article as: E.J. Kim and Z. Zhu, Variance function estimation of a one-dimensional nonstationary process. Journal of the Korean Statistical Society (2019), https://doi.org/10.1016/j.jkss.2019.01.001.

Table 1
Bandwidth selections for estimating (a) sine and (b) step $\sigma(\cdot)$ functions. Oracle bandwidths, λ^O, are defined to provide the minimum DMSE for difference-based and likelihood-based methods; bandwidth selections using our method, $\lambda*$, result in larger values than the oracle bandwidths. Bandwidths, marked *Levine*, are derived assuming the underlying process is independent (Levine, 2006). In parentheses are the variability.

n	Bandwidth	(a) Sine			(b) Step		
	Methods	$\theta = 0.1$	$\theta = 0.01$	indep.	$\theta = 0.1$	$\theta = 0.01$	indep.
100	Diff-λ^O	0.203	0.206	0.209	0.218	0.222	0.229
		(.054)	(.059)	(.052)	(.071)	(.084)	(.076)
	Diff-$\lambda*$	0.262	0.281	0.266	0.405	0.415	0.434
		(.074)	(.079)	(.069)	(.126)	(.087)	(.074)
	Levine	0.356	0.455	0.420	0.360	0.467	0.418
		(.297)	(.274)	(.281)	(.304)	(.267)	(.289)
	Like-λ^O	0.165	0.168	0.154	0.137	0.138	0.133
		(.054)	(.055)	(.033)	(.032)	(.030)	(.030)
200	Diff-λ^O	0.170	0.171	0.177	0.191	0.185	0.203
		(.034)	(.037)	(.046)	(.050)	(.060)	(.066)
	Diff-$\lambda*$	0.240	0.218	0.190	0.381	0.336	0.289
		(.090)	(.108)	(.119)	(.126)	(.143)	(.163)
	Levine	0.234	0.380	0.347	0.248	0.369	0.334
		(.248)	(.224)	(.229)	(.249)	(.230)	(.217)
	Like-λ^O	0.131	0.129	0.127	0.113	0.113	0.112
		(.034)	(.028)	(.021)	(.025)	(.024)	(.023)
500	Diff-λ^O	0.140	0.141	0.154	0.154	0.152	0.158
		(.027)	(.031)	(.037)	(.042)	(.042)	(.047)
	Diff-$\lambda*$	0.217	0.205	0.180	0.357	0.329	0.260
		(.107)	(.117)	(.111)	(.143)	(.147)	(.159)
	Levine	0.186	0.256	0.232	0.192	0.264	0.240
		(.186)	(.164)	(.165)	(.193)	(.152)	(.166)
	Like-λ^O	0.098	0.098	0.100	0.091	0.090	0.094
		(.016)	(.016)	(.016)	(.019)	(.016)	(.017)
1000	Diff-λ^O	0.120	0.121	0.133	0.131	0.125	0.148
		(.026)	(.026)	(.023)	(.033)	(.033)	(.038)
	Diff-$\lambda*$	0.209	0.186	0.170	0.329	0.300	0.255
		(.121)	(.117)	(.109)	(.159)	(.157)	(.165)
	Levine	0.180	0.289	0.174	0.199	0.288	0.191
		(.155)	(.118)	(.094)	(.157)	(.123)	(.092)
	Like-λ^O	0.086	0.084	0.086	0.078	0.076	0.078
		(.013)	(.011)	(.013)	(.015)	(.013)	(.014)

6. Summary

We developed a nonparametric variance function estimator for a one-dimensional nonstationary process using a difference filter. We assumed that the error process is additive and second-order stationary after normalizing the variability. We defined local variogram and derived infill asymptotic properties of the local variogram estimator dependent on the relative smoothness of the mean and variance functions and the mean square differentiability of the process.

We have shown through a simulation study that the difference-based estimation overall has a smaller DMSE than a likelihood-based approach. In nonparametric regression, the boundary bias can be easily fixed by adjusting the objective function, whereas the likelihood-based method adds generalized estimating equations to adjust the effect. Another contrast between the two approaches is in computing time. The difference-based method may require a matrix inversion for the correlation parameter estimation, when using a likelihood-based estimation, whereas the likelihood-based local variance estimation proposes an ad hoc bandwidth selection that requires a matrix inversion for every simulated process generation. The difference-based approach reduces the computing time by $O(n^{-2})$ to that of the likelihood-based method.

We extend the difference-based variance function estimation of a one-dimensional error process to a two-dimensional nonstationary random field (Kim & Zhu, 2017). The configurations for a difference filter in 2-D space are manifold with considerations for direction, weight, and the spatial distribution of observational points. The area of applications is also extensive for two-dimensional nonstationary random fields, and the feasibility is examined in a simpler, one-dimensional nonstationary process scenario in this paper.

Appendix. Technical details

Here is the Taylor expansion of the local variance of an $\frac{h}{n}$-lagged nonstationary process with smooth mean and variance functions. It details Eq. (3) to derive the local variogram (4) as the main term in the expansion.

$$var\left(Z\left(s-\frac{h}{2n}\right)-Z\left(s+\frac{h}{2n}\right)\right)$$

Please cite this article as: E.J. Kim and Z. Zhu, Variance function estimation of a one-dimensional nonstationary process. Journal of the Korean Statistical Society (2019), https://doi.org/10.1016/j.jkss.2019.01.001.

$$
\begin{aligned}
&= 2\left(\sigma^2(s) + \frac{\sigma^{2(2)}(s)}{2!}\left(\frac{h}{2n}\right)^2 + \frac{\sigma^{2(4)}(s)}{4!}\left(\frac{h}{2n}\right)^4 + o\left(\left(\frac{h}{2n}\right)^5\right)\right) \\
&\quad - 2\rho_h \sum_{k=0}^{p}\left\{\left(\frac{\sigma^{(k)}(s)}{k!}\right)^2\left(\frac{h}{2n}\right)^{2k}(-1)^k + 2\sum_{i+j=2k,\, i\neq j}\frac{\sigma^{(i)}(s)}{i!}\frac{\sigma^{(j)}(s)}{j!}\left(\frac{h}{2n}\right)^{2k}\right\} \\
&= 2(1-\rho_h)\left\{\sigma^2(s) + \frac{\sigma^{2(2)}(s)}{2!}\left(\frac{h}{2n}\right)^2 + \frac{\sigma^{2(4)}(s)}{4!}\left(\frac{h}{2n}\right)^4 + o\left(\left(\frac{h}{2n}\right)^5\right)\right\} \\
&\quad + \rho_h\left[\frac{\left(\sigma^{2(1)}(s)\right)^2}{\sigma^2(s)}\left(\frac{h}{2n}\right)^2\right. \\
&\quad \left. + \left\{\frac{\left(\sigma^{2(1)}(s)\right)^4}{32\left(\sigma^2(s)\right)^3} + \frac{\left(\sigma^{2(1)}(s)\right)^2}{8\left(\sigma^2(s)\right)^2} - \frac{3\left(\sigma^{2(2)}(s)\right)^2}{8\sigma^2(s)} + \frac{\sigma^{2(1)}(s)\sigma^{2(3)}(s)}{6\sigma^2(s)}\right\}\left(\frac{h}{2n}\right)^4\right] \\
&= 2\sigma^2(s)(1-\rho_h) + \left\{\sigma^{2(2)}(s)(1-\rho_h) + \frac{\left(\sigma^{2(1)}(s)\right)^2}{\sigma^2(s)}\rho_h\right\}\left(\frac{h}{2n}\right)^2 + o\left(\left(\frac{h}{2n}\right)^3\right). \qquad (19)
\end{aligned}
$$

$\delta_{i,h} \le c_\mu^2 (h/n)^q$. Under the condition that $\mu(\cdot) \in \Lambda_q(c_f)$ and $q \ge 0$, the Taylor expansion of $\delta_{i,h}$ about location s when $q \ge 1$ is:

$$
\begin{aligned}
\delta_{i,h} &= \sum_{j=1}^{\lfloor q\rfloor}\frac{\mu_s^{(j)}}{j!}\left\{(s_i - s)^j - (s_{i+h} - s)^j\right\} + O\left(|s_i - s|^q + |s_{i+h} - s|^q\right) \\
&= -\frac{h}{n}\sum_{j=1}^{\lfloor q\rfloor}\frac{\mu_s^{(j)}}{j!}\sum_{a=0}^{j-1}(s_i - s)^a(s_{i+h} - s)^{j-1-a} + O(|s_i - s|^q + |s_{i+h} - s|^q); \qquad (20)
\end{aligned}
$$

and when $0 \le q < 1$, it is:

$$
\delta_{i,h} = c\left(\frac{i}{n}\right)^q - c\left(\frac{i+h}{n}\right)^q = O\left(n^{-q}\right). \qquad (21)
$$

A Taylor expansion of $g_{i,h}$ about location s is:

$$
\begin{aligned}
\frac{1}{2}g_{i,h} &= (1-\rho_h)\left[\sigma_s^2 + \sigma_s\sum_{j=1}^{\lfloor\beta\rfloor}\frac{\sigma_s^{(j)}}{j!}\left\{(s_i - s)^j + (s_{i+h} - s)^j\right\}\right] + O(|s_i - s|^\beta) \\
&\quad + \sum_{l=1}^{\lfloor\beta/2\rfloor}\left(\frac{\sigma_s^{(l)}}{l!}\right)^2\left\{(s_i - s)^{2l} + (s_{i+h} - s)^{2l} - \rho_h(s_i - s)^l(s_{i+h} - s)^l\right\} \\
&\quad + \sum_{m=3}^{\lfloor\beta\rfloor}\sum_{j=1}^{m-1}\left[\frac{c_m\sigma_s^{(j)}\sigma_s^{(m-k)}}{m!}\left\{(s_i - s)^m + (s_{i+h} - s)^m\right\} - \rho_h\frac{\sigma_s^{(j)}\sigma_s^{(m-j)}}{j!(m-j)!}(s_i - s)^j(s_{i+h} - s)^{m-j}\right] \qquad (22)
\end{aligned}
$$

under the condition that $\sigma^2(\cdot) \in \Lambda_\beta^+$ and $\beta \ge 2$.

Remark 4. We detail Theorem 3.1 in the order we listed the results in (7).

A. Assume that $m > q$ and $m > \beta$, in other words the order of kernel is greater than the degree differentiability of both the mean and variance functions. Then, (A.i) when $\alpha < 1$ and $\frac{\alpha+1}{2} < q \le 1$, the bias is $O\left(n^{-\alpha-1}\right)$; (A.ii) when $\alpha < 1$ and $2q \le \alpha + 1/2$, the bias is $O\left(n^{-2q}\right)$; and (A.iii) when $\alpha \ge 1$ and $q \ge 1$, the bias is $O\left(n^{-2}\right)$.

B. Assume that $q < m \le \beta$ and that $\lambda = O(n^{-x})$ where $0 < x < 1$. Then $O(n^{-\alpha}\lambda^m)$ is the order of bias in the following three settings: (B.i) $q \ge 1$, $\alpha \le 1$, and $x < 1/m$; (B.ii) $q \ge 1$, $\alpha \ge 1$, and $x < (2-\alpha)/m$; and (B.iii) $\alpha < 1$, $2q < \alpha + 1$, and $x < (2q - \alpha)/m$. The remaining scenarios should refer to Case A.

C. Assume that $m \le q$ irrespective of β and that $\lambda = O(n^{-x})$ where $0 < x < 1$. Then the bias is $O(\lambda^m)$ in the following three settings: (C.i) $q \ge 1$, $\alpha \ge 1$, and $2/m > x$; (C.ii) $q < \min(1, \frac{\alpha+1}{2})$, and $x < 2q/m$; (C.iii) $\alpha < 1$, $\alpha + 1 < 2q$ and $x < (\alpha + 1)/m$. The remaining scenarios should refer to Case A.

Please cite this article as: E.J. Kim and Z. Zhu, Variance function estimation of a one-dimensional nonstationary process. Journal of the Korean Statistical Society (2019), https://doi.org/10.1016/j.jkss.2019.01.001.

References

Anderes, E. B., & Stein, M. L. (2011). Local likelihood estimation for nonstationary random fields. *Journal of Multivariate Analysis*, *102*, 506–520.
Brown, L. D., & Levine, M. (2007). Variance estimation in nonparametric regression via the difference sequence method. *The Annals of Statistics*, *35*(5), 2219–2232.
Cai, T. T., & Wang, L. (2008). Adaptive variance function estimation in heteroscedastic nonparametric regression. *The Annals of Statistics*, *36*(5), 2025–2054.
Gasser, T., Müller, H. -G., & Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *Journal of Royal Statistical Society. Series B*, *47*(2), 238–252.
Gasser, T., Sroka, L., & Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika*, *73*(3), pp. 625–633.
Hall, P., & Carroll, R. J. (1989). Variance function estimation in regression: the effect of estimating the mean. *Journal of Royal Statistical Society. Series B*, *51*(1), 3–14.
Hall, P., Kay, J. W., & Titterington, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika*, *77*, 521–528.
Hall, P., Kay, J. W., & Titterington, D. M. (1991). On estimation of noise variance in two-dimensional signal processing. *Advanced Applied Probability*, *23*, 476–495.
Kim, E. J., & Zhu, Z. (2017). Estimating a variance function of a nonstationary process. In D. A. Griffith, Y. Chun, & D. J. Dean (Eds.), *Advances in geocomputation*. Springer.
Levine, M. (2006). Bandwidth selection for a class of difference-based variance estimators in the nonparametric regression: A possible approach. *Computational Statistics & Data Analysis*, *50*(12), 3405–3431.
Matheron, G. (1962). Traité de Géostatistique Appliquée, Tome I. *Mémoires du Bureau de Recherches Géologiques et Minieres*: *vol. 14*, (p. 333). Mémoires du Bureau de Recherches Géologiques et Minières, Paris.
Neumann, J. V., Kent, R. H., Bellinson, H. R., & Hart, B. I. (1941). The mean square successive difference. *The Annals of Mathematical Statistics*, *12*(2), 153–162.
Opsomer, J., Wang, Y., & Yang, Y. (2001). Nonparametric regression with correlated errors. *Statistical Science*, *16*(2), 134–153.
Seifert, B., Gasser, T., & Wolf, A. (1993). Nonparametric estimation of residual variance revisited. *Biometrika*, *80*(2), 373–383.
Stein, M. L. (1999). *Interpolation of spatial data: Some theory for kriging*. Springer.
Tsybakov, A. (2009). *Introduction to nonparametric estimation*. Springer.
Wang, L., Brown, L. D., Cai, T. T., & Levine, M. (2008). Effect of mean on variance function estimation in nonparametric regression. *The Annals of Statistics*, *36*(2), 646–664.

Please cite this article as: E.J. Kim and Z. Zhu, Variance function estimation of a one-dimensional nonstationary process. Journal of the Korean Statistical Society (2019), https://doi.org/10.1016/j.jkss.2019.01.001.