EMPIRICAL-BIAS BANDWIDTHS FOR SPATIAL LOCAL POLYNOMIAL REGRESSION WITH CORRELATED ERRORS

TORGNY LINDSTRÖM

Preprints in Mathematical Sciences
2004:11

LUND INSTITUTE OF TECHNOLOGY
Lund University

Centre for Mathematical Sciences
Mathematical Statistics

CENTRUM SCIENTIARUM MATHEMATICARUM

Empirical-Bias Bandwidths for Spatial Local Polynomial Regression with Correlated Errors

Torgny Lindström

Abstract

The empirical-bias bandwidth selector (EBBS) is a method for data-driven selection of bandwidths for local polynomial regression. It is a bandwidth selection method for estimation of the mean-function and its partial derivatives of any order as well as for estimation of the variance-function. Moreover EBBS allows for univariate as well as multivariate predictor variables.

In this paper we introduce the empirical-bias bandwidth selector, $\mathrm{EBBS}_{\mathrm{dep}}$. This estimation procedure is adjusted to allow for dependent errors and selection of diagonal or full bandwidth matrices for estimation of the mean-function or one of its partial derivatives as well as for estimation of the variance-function. Asymptotic results for the conditional bias of the first order partial derivative estimates are given for the local quadratic regression case.

A simulation study is performed to compare the adjusted and the original version of EBBS with theoretical results for a few cases displaying varying degrees of positive correlation.

Key words

Derivative estimation; heteroscedasticity; local bandwidth selection; surface fitting; variance-function estimation.

1 Introduction

Nonparametric regression methods are useful for visualizing or estimating statistical characteristics of spatial data like e.g. the mean-function or the variance-function. The advantage with such methods is that no prior assumptions about the parametric shape of the function in question need to be imposed upon the model.

However, in order to correctly enhance the features of interest, it is important to have good methods for bandwidth selection, so that the amount of smoothing can be controlled. These must be chosen depending on the characteristics of the data and on the features of interest. Using bandwidths that are too small would lead to unnecessarily noisy estimates which might hide true features of the data or even show false peaks. On the other hand, too large bandwidths would cause true peaks to be over-smoothed resulting in an estimate that might miss important features.

For data showing varying degrees of spatial trend or variability depending on location it might be advantageous to use local smoothing parameters. By considering local smoothing parameter selection where one or more parameters need to be selected for each location, it is clear that automatic selection methods are needed. Even for global bandwidths a "trial and error" approach would not be possible other than in very low dimensions. Such a task is inevitably time consuming and would project the analysts own perception of the true underlying function onto the fit. Also for higher dimensions the problem of visualizing the fits is another obstacle for this approach.

A number of different schemes for bandwidth selection have been proposed in the literature, most of which are based on the assumption of independent errors. Since spatial data often are correlated, available criteria might not be suitable for bandwidth selection. Recently, Francisco-Fernández and Opsomer (2003) suggested a selection approach for global bandwidths based on a correlation-adjusted generalized cross-validation (GCV) criterion.

In this paper we consider a method for local bandwidth selection that minimizes an estimate of the mean squared error (MSE) of the local polynomial estimator. Utilizing an empirical estimate of the bias and the known expression for the variance, an estimate of the MSE is produced by using the decomposition into a variance term and a squared bias term. Based on the empirical-bias bandwidth selector (EBBS), proposed by Ruppert (1997), we introduce the empirical-bias bandwidth selector adjusted to allow for bandwidth matrices and dependent errors, $\mathrm{EBBS}_{\mathrm{dep}}$. The focus of the presentation is on the bivariate case. Though this method is easily generalized to higher dimensions in theory, the implementation of the algorithm is increasingly extensive because of numerical considerations.

Section 2 describes the model together with a brief overview of bivariate local polynomial regression. Also included is a short review of EBBS and a detailed presentation of the $\mathrm{EBBS}_{\mathrm{dep}}$ procedure. In Section 3 simulation studies are used for illustrations and for evaluation of the procedure. The results are summarized in Section 4. Asymptotic results for the conditional bias of first order derivative estimates using local quadratic regression are given in the Appendix.

2 Empirical-Bias Bandwidths for Spatial Data with Dependent Errors

2.1 The Model

Let $\{(\mathbf{X}_i, Y_i), i = 1, \ldots, n\}$ be a random sample, where for each i, $\mathbf{X}_i = (X_{i,1}, X_{i,2})^\mathsf{T}$ is a bivariate vector describing the location of the response Y_i. These are assumed to satisfy the model,

$$Y_i = m(\mathbf{X}_i) + v^{1/2}(\mathbf{X}_i)\varepsilon_i, \;\; i = 1, \ldots, n, \tag{1}$$

where the mean and the variance are specified by the smooth functions $m(\mathbf{x}) = \mathsf{E}(Y|\mathbf{X} = \mathbf{x})$ and $v(\mathbf{x}) = \mathsf{Var}(Y|\mathbf{X} = \mathbf{x})$ respectively. The random errors, $\varepsilon_1, \ldots, \varepsilon_n$, are assumed to be random variables with $\mathsf{E}(\varepsilon_i|\mathbf{X}_i = \mathbf{x}_i) = 0$, $\mathsf{Var}(\varepsilon_i|\mathbf{X}_i = \mathbf{x}_i) = 1$ and $\mathsf{Cov}(\varepsilon_i, \varepsilon_j|\mathbf{X}_i = \mathbf{x}_i, \mathbf{X}_j = \mathbf{x}_j) = \rho(||\mathbf{x}_i - \mathbf{x}_j||)$. Let $\boldsymbol{\Sigma}$ denote the covariance matrix of $Y_1, \ldots, Y_n$ given the locations $\{\mathbf{X}_i\}$ which can be either fixed or random with a common density.

All moments in this paper are meant to be conditional on $\mathbf{X}_1, \ldots, \mathbf{X}_n$, i.e. expressions for bias, variance, etc are all conditional on the locations. We will however follow the convention in kernel regression and omit the conditional notation.

2.2 Local Polynomial Regression

Following is a brief overview of bivariate local polynomial estimation of $m(\cdot)$ or one of its partial derivatives denoted by,

$$m^{(k_1,k_2)}(\cdot) = \frac{\partial^K}{\partial x_1^{k_1} \partial x_2^{k_2}} \, m(\cdot), \tag{2}$$

where $k_j \geq 0$, for $j = 1, 2$, and $K = k_1 + k_2$.

Consider estimation of $m(\mathbf{x})$ which, by the smoothness assumption, can be approximated locally by fitting a polynomial to $\{(\mathbf{X}_i, Y_i)\}$, $i = 1, \ldots, n$, by weighted least squares. The general shape of the weight-function defining the local structure is given by the kernel function $K(\cdot)$, a bivariate nonnegative function satisfying $\int K(\mathbf{u})d\mathbf{u} = 1$ and $\int \mathbf{u}K(\mathbf{u})d\mathbf{u} = 0$. The kernel function is usually a unimodal spherically contoured density function or a product of univariate density functions. The regression weight given to location $\mathbf{X}_i$ for estimation at $\mathbf{x}$ is obtained from

$$K_H(\mathbf{X}_i - \mathbf{x}) = |\mathbf{H}|^{-1} K(\mathbf{H}^{-1}(\mathbf{X}_i - \mathbf{x})), \tag{3}$$

where $\mathbf{H}$ is a 2×2 symmetric positive definite matrix called the bandwidth matrix. This matrix controls the shape and size of the local neighbourhood. Assuming a kernel function with spherical contours, the diagonal elements of $\mathbf{H}$ is the scaling for the respective components of $\mathbf{x}$, while the off-diagonal elements introduce rotation of the elliptically shaped contours.

The local polynomial estimator of $m^{(k_1,k_2)}(\mathbf{x})$ is found by minimization of,

$$\sum_{i=1}^{n}\{Y_i - P_p(\mathbf{X}_i;\boldsymbol{\beta}(\mathbf{x}))\}^2 K_H(\mathbf{X}_i - \mathbf{x}), \tag{4}$$

with respect to $\boldsymbol{\beta}(\mathbf{x})$. Here $P_p(\mathbf{X}_i;\boldsymbol{\beta}(\mathbf{x}))$ is a p-degree polynomial in $\mathbf{X}_i$,

$$P_p(\mathbf{X}_i;\boldsymbol{\beta}(\mathbf{x})) = \sum_{K=0}^{p}\sum_{k_1+k_2=K}\beta_{k_1,k_2}(\mathbf{x})\prod_{j=1}^{2}(X_{i,j} - x_j)^{k_j}, \tag{5}$$

and $\boldsymbol{\beta}(\mathbf{x}) = \{\beta_{k_1,k_2}(\mathbf{x}) : k_1 + k_2 = K;\ K = 0,\ldots,p\}$ is the vector of polynomial coefficients.

The estimator of $m^{(k_1,k_2)}(\mathbf{x})$ is given by

$$\hat{m}_{k_1,k_2}(\mathbf{x};\mathbf{H},p) = D_K\,\hat{\beta}_{k_1,k_2}(\mathbf{x}), \tag{6}$$

where $D_K = \prod_{j=1}^{2}(k_j!)$ and $\hat{\beta}_{k_1,k_2}(\mathbf{x})$ is the appropriate component, say the l:th, of the vector

$$\hat{\boldsymbol{\beta}}(\mathbf{x}) = (\mathbf{X}_{p,x}^{\mathsf{T}}\mathbf{W}_{H,x}\mathbf{X}_{p,x})^{-1}\mathbf{X}_{p,x}^{\mathsf{T}}\mathbf{W}_{H,x}\mathbf{Y}, \tag{7}$$

where $\mathbf{W}_{H,x} = \operatorname{diag}\{K_H(\mathbf{X}_1 - \mathbf{x}),\ldots,K_H(\mathbf{X}_n - \mathbf{x})\}$ is the $n \times n$ diagonal matrix of weights, $\mathbf{Y} = (Y_1,\ldots,Y_n)^{\mathsf{T}}$ and $\mathbf{X}_{p,x}$ denotes the bivariate p-degree design matrix for regression at location $\mathbf{x}$. Estimation of $m^{(k_1,k_2)}(\cdot)$ at some set of locations, $\mathbf{x}'_1,\ldots,\mathbf{x}'_m$, is given by $\hat{m}_{k_1,k_2}(\cdot;\mathbf{H},p) = D_K\,\hat{\boldsymbol{\beta}}_{k_1,k_2}(\cdot)$. Here the $m \times 1$ vector of regression coefficients $\hat{\boldsymbol{\beta}}_{k_1,k_2}(\cdot)$ is,

$$\hat{\boldsymbol{\beta}}_{k_1,k_2}(\cdot) = \mathbf{S}(l)\mathbf{Y}, \tag{8}$$

where $\mathbf{S}(l)$ is the $m \times n$ matrix for estimation of the (k_1,k_2):th partial derivative, with components

$$\mathbf{S}(l)_{i,j} = \mathbf{e}_l^{\mathsf{T}}(\mathbf{X}_{p,x'_i}^{\mathsf{T}}\mathbf{W}_{H,x'_i}\mathbf{X}_{p,x'_i})^{-1}\mathbf{X}_{p,x'_i}^{\mathsf{T}}\mathbf{W}_{H,x'_i}\mathbf{e}_j \tag{9}$$

and $\mathbf{e}_l$ is a column vector of zeros except for the value at position l which is one. Note that $\mathbf{e}_l$ and $\mathbf{e}_j$ above are of different lengths. In order to keep the notation short the lengths of such vectors will not be explicitly stated.

2.3 The Empirical-Bias Bandwidths Selector

The method for selection of local bandwidths is based on the empirical-bias bandwidth selector proposed by Ruppert (1997). Below follows a short review of EBBS.

The object of this method is to select a sequence of scalar bandwidths, $h(\mathbf{x})$, which in the independent case minimizes the estimate of the mean squared error (MSE) of $\hat{m}_{k_1,k_2}(\mathbf{x}; h, p)$ given by,

$$\begin{aligned}\mathsf{E}\{\hat{m}_{k_1,k_2}(\mathbf{x}; h, p) - m^{(k_1,k_2)}(\mathbf{x})\}^2 = \mathsf{Var}\{\hat{m}_{k_1,k_2}(\mathbf{x}; h, p)\} \\ + [\mathsf{E}\{\hat{m}_{k_1,k_2}(\mathbf{x}; h, p)\} - m^{(k_1,k_2)}(\mathbf{x})]^2.\end{aligned} \tag{10}$$

Estimation of the MSE is given by combining separate estimates of the bias and the variance term in (10).

The variance term is straightforward to obtain and utilizes weighted least squares theory,

$$\mathsf{Var}(\hat{\boldsymbol{\beta}}(\mathbf{x})) = (\mathbf{X}_{p,x}^{\mathsf{T}}\mathbf{W}_{h,x}\mathbf{X}_{p,x})^{-1}(\mathbf{X}_{p,x}^{\mathsf{T}}\mathbf{W}_{h,x}\boldsymbol{\Sigma}\mathbf{W}_{h,x}\mathbf{X}_{p,x})(\mathbf{X}_{p,x}^{\mathsf{T}}\mathbf{W}_{h,x}\mathbf{X}_{p,x})^{-1}. \tag{11}$$

Assume that the variance-function is approximately constant in a local neighbourhood of $\mathbf{x}$ so that $v(\mathbf{X}_j) \approx v(\mathbf{x})$ for $\mathbf{X}_j$ such that $K_h(\mathbf{X}_j - \mathbf{x}) \neq 0$. Now, using the assumption that the errors are independent, the variance matrix $\boldsymbol{\Sigma}$ in (11) can be replaced by the identity matrix times $v(\mathbf{x})$. Let $\hat{\beta}_{k_1,k_2}(\mathbf{x})$ be the l:th component of $\hat{\boldsymbol{\beta}}(\mathbf{x})$. Then the variance of the estimate of the (k_1, k_2)):th partial derivative of $m(\cdot)$ at $\mathbf{x}$ is approximately given by,

$$D_K^2\, v(\mathbf{x})\, \mathbf{e}_l^{\mathsf{T}}(\mathbf{X}_{p,x}^{\mathsf{T}}\mathbf{W}_{h,x}\mathbf{X}_{p,x})^{-1}(\mathbf{X}_{p,x}^{\mathsf{T}}\mathbf{W}_{h,x}\mathbf{W}_{h,x}\mathbf{X}_{p,x})(\mathbf{X}_{p,x}^{\mathsf{T}}\mathbf{W}_{h,x}\mathbf{X}_{p,x})^{-1}\mathbf{e}_l. \tag{12}$$

Now, to calculate the estimate of the variance is just a matter of plugging an estimate of the variance-function into (12). In EBBS the local polynomial variance-function estimator of Ruppert et al. (1997) is used for this purpose.

The bias term in (10) is estimated empirically. Consider estimation of the bias for a given bandwidth h_0 at location $\mathbf{x}$. By calculating $\hat{m}_{k_1,k_2}(\mathbf{x}; h, p)$ on a grid of h-values in a neighbourhood of h_0 a polynomial is fitted to the $\{h, \hat{m}_{k_1,k_2}(\mathbf{x}; h, p)\}$ "data". The form of the polynomial is based on the asymptotic properties of $\hat{m}_{k_1,k_2}(\mathbf{x}; h, p)$. From this polynomial fit an estimate of the bias can be computed. For a thorough description of the EBBS algorithm, see Ruppert (1997).

2.4 EBBS for Dependent Errors, EBBS$_{\text{dep}}$

In this section we introduce EBBS for selection of local bivariate bandwidth matrices for models with possibly dependent errors, EBBS$_{\text{dep}}$. It is an estimation procedure involving an iterative sequence for estimation of $\boldsymbol{\Sigma}$ similar to the one proposed by Opsomer et al. (1999). The following steps present an overview of EBBS$_{\text{dep}}$ for estimation of the the mean-function or one of its derivatives.

1. Initially, obtain a mean-function estimate $\hat{m}_{0,0}(\mathbf{X}_i; \mathbf{H}, p)$, $i = 1, \ldots, n$, at the locations for the observations, using polynomials of degree $p \geq 2$ and constant spans for the bandwidths. Construct the residuals η_i, $i = 1, \ldots, n$.

   The span $0 < q \leq 1$ at location $\mathbf{x}$ for $\mathbf{H}_{jj}(\mathbf{x})$, $j = 1, 2$ is the smallest value of $\mathbf{H}_{jj}(\mathbf{x})$ such that at least nq of $X_{1,j}, \ldots, X_{n,j}$ is within distance $\mathbf{H}_{jj}(\mathbf{x})$ of $\mathbf{x}$.

2. Make an initial estimate of $v(\cdot)$ at $\mathbf{X}_1, \ldots, \mathbf{X}_n$ using local polynomial smoothing of the squared residuals. The bandwidth matrices used here are based on the initial selection procedure described in Section 2.4.4.

3. Obtain the standardized residuals $\tilde{\eta}_i = \eta_i / (\hat{v}(\mathbf{X}_i)(1 + \Delta_i))^{1/2}$. The expression for $\tilde{\eta}_i$ is motivated in Section 2.4.5.

4. Use the $\tilde{\eta}_i$'s to estimate the correlation-function $\rho(\cdot)$.

5. Calculate $\widehat{\boldsymbol{\Sigma}}$ where $\widehat{\boldsymbol{\Sigma}}_{ij} = \hat{\rho}(||\mathbf{X}_i - \mathbf{X}_j||)\hat{v}(\mathbf{X}_i)^{1/2}\hat{v}(\mathbf{X}_j)^{1/2}$.

6. Update the estimate of $m(\cdot)$ using bandwidths based on $\widehat{\boldsymbol{\Sigma}}$ and update the residuals, $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^{\mathsf{T}}$.

7. Reestimate $\hat{\mathbf{v}} = (\hat{v}(\mathbf{X}_1), \ldots, \hat{v}(\mathbf{X}_n))^{\mathsf{T}}$, $\tilde{\boldsymbol{\eta}}$ and $\hat{\rho}(\cdot)$.

8. Update $\widehat{\boldsymbol{\Sigma}}$ and obtain the estimate of $m^{(k_1,k_2)}(\cdot)$.

In the estimation procedure steps 5–7 can be iterated in order to improve the estimate of $\boldsymbol{\Sigma}$.

The algorithm for bandwidth selection in EBBS$_{\text{dep}}$ is built on the same principle as EBBS though the support for bandwidth matrices and for dependent errors conveys differences in the details. Therefore we have chosen to give a fairly detailed description of the procedure. The notation from Ruppert (1997) is preserved as much as possible in order to facilitate comparisons.

2.4.1 Preliminaries

Estimation is performed on a grid $G_x = \{\mathbf{x}'_j\}$, $j = 1, \ldots, m$. Often G_x is regularly spaced even if any kind of grid is feasible.

For some fixed k_1, k_2 and p and for each $\mathbf{x}'_l \in G_x$, $\widehat{\text{MSE}}(\hat{m}_{k_1,k_2}(\mathbf{x}'_l; \mathbf{H}, p))$ is calculated for a set of symmetric positive definite matrices $\mathbf{H}$, where

$$\mathbf{H} = \begin{pmatrix} h_1 & h_{12} \\ h_{12} & h_2 \end{pmatrix}. \tag{13}$$

In the EBBS$_{\text{dep}}$ software the range of valid bandwidths is chosen by the user by specifying upper and lower bounds for spans for each of h_1 and h_2. The range for h_{12} is then given by $h_{12}^2 < h_1 h_2$ to ensure positive definiteness of $\mathbf{H}$.

2.4.2 Bias Estimation

In the following we describe the algorithm for estimating the bias of $\hat{m}_{k_1,k_2}(\mathbf{x};\mathbf{H},p)$ as a function of $\mathbf{H}$. Let $\mathbf{x}$ be fixed and consider bias estimation for $\mathbf{H}_0$ consisting of $(h_{1,0},h_{2,0},h_{12,0})$ in the user supplied range of triples $G_U(\mathbf{x}) = \{(h_1,h_2,h_{12})\}$ for location $\mathbf{x}$. Construct a new grid of triples, $G_{H_0}(\mathbf{x})$, in a neighbourhood of $(h_{1,0},h_{2,0},h_{12,0})$ and calculate $\hat{m}_{k_1,k_2}(\mathbf{x};\mathbf{H}_{0,j},p)$ for all $\mathbf{H}_{0,j}$ consisting of triples in $G_{H_0}(\mathbf{x})$.

For a fixed integer $t \geq 1$ fit by ordinary least squares the polynomial,

$$c_0(\mathbf{x}) + c_{p-K+1}(\mathbf{x})\,\mathrm{tr}(\mathbf{H}^2)^{(p-K+1)/2} + \ldots + c_{p-K+t}(\mathbf{x})\,\mathrm{tr}(\mathbf{H}^2)^{(p-K+t)/2}, \qquad (14)$$

to the "data" $\{\mathrm{tr}(\mathbf{H}_{0,j}^2), \hat{m}_{k_1,k_2}(\mathbf{x};\mathbf{H}_{0,j},p)\}$. The form of the polynomial (14) is motivated by the results in Ruppert and Wand (1994), giving the asymptotic conditional bias of $\hat{m}_{0,0}(\mathbf{x};\mathbf{H},p)$ for general p and $t = 1$. Hence the terms after the first one in (14) are describing the (asymptotic) bias.

The results for derivative estimation in multivariate regression can be derived following the methodology for estimation of the mean. Results for first order derivative estimation by quadratic polynomials are given in Theorems 2 and 3 in the Appendix. This special case is motivated by the application concerning evaluation of lidar measurements considered by Lindström et al. (2004).

Also, note that the results of Ruppert and Wand (1994) consider the independent case. Liu (2001, Theorem 2.1, page 34) showed corresponding results for the variance for correlated data and $p = 1$. Note that the asymptotic conditional bias is not affected by correlation.

The bias of $\hat{m}_{k_1,k_2}(\mathbf{x};\mathbf{H}_0,p)$ is thus estimated by,

$$\hat{c}_{p-K+1}(\mathbf{x})\,\mathrm{tr}(\mathbf{H}_0^2)^{(p-K+1)/2} + \ldots + \hat{c}_{p-K+t}(\mathbf{x})\,\mathrm{tr}(\mathbf{H}_0^2)^{(p-K+t)/2}. \qquad (15)$$

As noted by Ruppert (1997), for $p-K$ even it is advisable to use $t \geq 2$ since $c_j = 0$ for odd values of j except at the boundary.

2.4.3 Variance Estimation

Estimating the variance of $\hat{m}_{k_1,k_2}(\mathbf{x};\mathbf{H},p) = D_K\hat{\beta}_{k_1,k_2}(\mathbf{x})$ is straightforward since $\hat{\beta}_{k_1,k_2}(\mathbf{x})$ is a weighted least squares estimate, see e.g Fan and Gijbels (1996).

$$\begin{aligned}\mathsf{Var}\{\hat{m}_{k_1,k_2}(\mathbf{x};\mathbf{H},p)\} = {} & D_K^2\mathbf{e}_l^{\mathsf{T}}(\mathbf{X}_{p,x}^{\mathsf{T}}\mathbf{W}_{H,x}\mathbf{X}_{p,x})^{-1} \\ & \times\, (\mathbf{X}_{p,x}^{\mathsf{T}}\mathbf{W}_{H,x}\boldsymbol{\Sigma}\mathbf{W}_{H,x}\mathbf{X}_{p,x}) \\ & \times\, (\mathbf{X}_{p,x}^{\mathsf{T}}\mathbf{W}_{H,x}\mathbf{X}_{p,x})^{-1}\mathbf{e}_l.\end{aligned} \qquad (16)$$

Recall that $\hat{\beta}_{k_1,k_2}(\mathbf{x})$ is the l:th component of $\hat{\boldsymbol{\beta}}(\mathbf{x})$. Then the covariance matrix of $\hat{m}_{k_1,k_2}(\cdot;\mathbf{H},p)$ is given by

$$\mathsf{Cov}\{\hat{m}_{k_1,k_2}(\cdot;\mathbf{H},p)\} = D_K^2\mathbf{S}_{G_x}(l)\boldsymbol{\Sigma}\mathbf{S}_{G_x}(l)^{\mathsf{T}}, \qquad (17)$$

where $\mathbf{S}_{G_x}(l)$ is the $m \times n$ matrix for estimation of $\hat{m}_{k_1,k_2}(\cdot; \mathbf{H}, p)$ on the grid G_x, cf. (9). In this expression all quantities are known except $\boldsymbol{\Sigma}$ which needs to be estimated. This estimate is constructed from separate estimates of $\rho(\cdot)$ and $v(\cdot)$ in Section 2.4.5. Note that no approximation is made in the expression for the covariance matrix of $\hat{m}_{k_1,k_2}(\cdot; \mathbf{H}, p)$.

2.4.4 Variance-function Estimation

Variance-function estimation is based on the local polynomial variance-function estimator proposed by Ruppert et al. (1997). However, since they assume that the observations are independent, the method here is slightly different. Also the conditional bias and variance of the estimator are affected. Updated results for the variance-function estimator in this paper are given in the Appendix. The estimator is developed as follows.

Assume an initial estimate of the mean-function at the observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$, $\hat{\mathbf{m}} = (\hat{m}_1, \ldots, \hat{m}_n)^\mathsf{T}$ which is a linear smooth of the observations, $\hat{\mathbf{m}} = \mathbf{SY}$. Here $\mathbf{S}$ is a smoother matrix in which it is assumed that each row add up to one. Any kind of linear smoothers are feasible but we focus on the local polynomial smoother.

Calculate the residuals by making an initial smooth, using smoother matrix $\mathbf{S}_1$, of the observations, $\boldsymbol{\eta} = (\mathbf{I} - \mathbf{S}_1)\mathbf{Y}$. A variance-function estimate is then produced by making a second smooth of the squared residuals,

$$\tilde{\mathbf{v}} = \mathbf{S}_2(\boldsymbol{\eta} \odot \boldsymbol{\eta}), \tag{18}$$

where $\tilde{\mathbf{v}} = (\tilde{v}(\mathbf{X}_1), \ldots, \tilde{v}(\mathbf{X}_n))^\mathsf{T}$ and $\mathbf{A} \odot \mathbf{B}$ denotes the element-wise product between the equally sized matrices $\mathbf{A}$ and $\mathbf{B}$. Next the $\tilde{\mathbf{v}}$-estimate needs to be corrected for bias. Sources for bias include:

- Bias from the initial smooth, i.e the bias introduced in the mean-function estimate by local polynomial regression.
- Bias introduced by local polynomial regression for the (non-constant) mean-function of the squared residuals.
- The variability of $\hat{\mathbf{m}}$ renders bias to $\tilde{\mathbf{v}}$.
- The correlation between $\varepsilon_1, \ldots, \varepsilon_n$.

It is assumed that the bandwidths for the initial mean estimate are sufficiently small so that the bias from the first item in the list above is negligible, hence $\hat{\mathbf{m}}$ will capture most of the local structure in $\mathbf{Y}$. Thus it is reasonable to assume that $\hat{\rho}(\cdot)$ is of short range so the bias from the correlation between the ε_i's is small. This discussion relates to the identification problem that occurs in situations where nonparametric estimation techniques are used in models with correlated errors, cf. Opsomer et al. (2001). By correcting for the increased variability of $\hat{\mathbf{m}}$ we get the following estimator of $v(\cdot)$,

$$\hat{\mathbf{v}} = \mathbf{S}_2\{\boldsymbol{\eta} \odot \boldsymbol{\eta}\}/(\mathbf{1} + \mathbf{S}_2\boldsymbol{\Delta}), \tag{19}$$

where / denotes element-wise division and $\boldsymbol{\Delta}$ is the $n \times 1$ vector given by $\mathrm{Diag}\{\mathbf{S}_1\mathbf{S}_1^{\mathsf{T}} - 2\mathbf{S}_1\}$ where $\mathrm{Diag}\{\mathbf{A}\}$ is a column vector containing the diagonal components of the square matrix $\mathbf{A}$. Thus $\hat{\mathbf{v}}$ is unbiased if $v(\cdot)$ is constant, $\varepsilon_1, \ldots, \varepsilon_n$ are uncorrelated and $\hat{\mathbf{m}}$ is unbiased, cf. Ruppert et al. (1997, p. 263).

Also the bandwidths for smoothing the squared residuals are chosen by $\mathrm{EBBS}_{\mathrm{dep}}$. The procedure for estimation of the $\mathrm{MSE}(\tilde{v}(\mathbf{X}_i))$ differs depending on whether an estimate of $\boldsymbol{\Sigma}$ is available or not. The idea here is to allow for an iterative procedure where in the first iteration no user specified covariance-matrix is required. Instead the covariance-matrix of the squared residuals is estimated as follows. Assume that the errors, ε_i, come from a scale family so that for some $\varkappa > 0$,

$$\mathsf{Var}(v(\mathbf{X}_i)\varepsilon_i^2) = \mathsf{Var}(\varepsilon_i^2)v^2(\mathbf{X}_i) = \varkappa v^2(\mathbf{X}_i). \tag{20}$$

To estimate $\varkappa$ the squared residuals are sorted into $n_1 n_2$ rectangular shaped bins depending on the location. In each bin an estimate of $\varkappa$ is taken as the ratio of the estimated variance to the square of the estimated mean of the η_i^2's. The average of these ratios is then formed to get the estimate, $\hat{\varkappa}$.

When estimating the MSE of $\tilde{v}(\mathbf{X}_i; \mathbf{H}_0)$ for a given $\mathbf{X}_i$ and $\mathbf{H}_0$, the estimates of the variances of the squared residuals are given by $\hat{\varkappa}\,\tilde{v}^2(\mathbf{X}_i; \mathbf{H}_0)$. Further the η_i^2's are assumed to be uncorrelated, since at this stage in the estimation procedure the correlation-function is not yet estimated. They are also assumed to have constant variance in the local neighbourhood, i.e for all $\mathbf{x}$ such that $K_{H_0}(\mathbf{X_i} - \mathbf{x}) > 0$. The variance of $\tilde{v}(\mathbf{X}_i; \mathbf{H}_0)$ is then estimated by,

$$\mathsf{Var}\{\tilde{v}(\mathbf{X}_i; \mathbf{H}_0)\} = \hat{\varkappa}\,\tilde{v}^2(\mathbf{X}_i; \mathbf{H}_0)\, e_i \mathbf{S}_2 \mathbf{S}_2^{\mathsf{T}} e_i^{\mathsf{T}}. \tag{21}$$

In the case of an available estimate of $\boldsymbol{\Sigma}$ an approximate covariance matrix of the squared residuals is calculated as follows. Assume that the residuals are normally distributed with covariance matrix $\boldsymbol{\Sigma}^{(\eta)} = (\mathbf{I} - \mathbf{S}_1)\boldsymbol{\Sigma}(\mathbf{I} - \mathbf{S}_1)^{\mathsf{T}}$. Note that they have common mean zero since the bias of the initial mean-function estimate is negligible, as was assumed above. Then, according to Holmquist (1988), the covariance between the squared residuals η_i^2 and η_j^2 is given by

$$\boldsymbol{\Sigma}_{ij}^{(\eta^2)} = \mathsf{Cov}(\eta_i^2, \eta_j^2) = 2\left(\boldsymbol{\Sigma}_{ij}^{(\eta)}\right)^2. \tag{22}$$

Note that this also follows immediately from (36) if $\mathbf{S}_2 = \mathbf{I}_n$ since the numerator in this expression is the conditional covariance matrix of $\tilde{\mathbf{v}}$. Thus the covariance matrix of the variance estimates using smoother matrix $\mathbf{S}_2$ to smooth the squared residuals is

$$\mathsf{Cov}(\tilde{\mathbf{v}}) = \mathbf{S}_2 \boldsymbol{\Sigma}^{(\eta^2)} \mathbf{S}_2^{\mathsf{T}}, \tag{23}$$

where $\boldsymbol{\Sigma}^{(\eta^2)}$ is the covariance matrix of the squared residuals.

The bias in the initial mean-function estimate is an important factor which may significantly increase the variance-function estimate. Therefore it is important to ensure small bias by the use of small bandwidths for the initial mean-function estimate and by using polynomials of degree $p_1 > 1$. Ruppert et al. (1997) recommends $p_1 = 2$ or $p_1 = 3$. The increased variability of $\hat{\mathbf{m}}$ introduced by using small bandwidths is a minor problem since it is compensated for in the $\hat{\mathbf{v}}$ estimate.

If estimation of $v(\cdot)$ is the main objective replace estimation of $m^{(k_1,k_2)}(\cdot)$ in step 8 in the estimation procedure with $\hat{v}(\cdot)$.

2.4.5 Estimation of the Covariance Matrix

Now turn to the estimation of $\rho(\cdot)$ in steps 4 and 7. Since correlation-function estimation assumes a constant variance-function it is important to remove the effect of a varying $v(\cdot)$ from the residuals. Also, because of the bias in $\hat{\mathbf{v}}$ the variance of the $\tilde{\eta}_i$'s will generally not be one. Therefore, when estimating $\rho(\cdot)$, it is suggested to use covariance-function estimation. In this way $\hat{\rho}(0)$ can be used as a correcting factor for the general bias in $\hat{\mathbf{v}}$.

It seems reasonable to develop the procedure such that the estimate of $\boldsymbol{\Sigma}$ is unbiased if $v(\cdot)$ is constant and the errors are iid. Assuming that the bias of the initial smooth is negligible, $\mathsf{Var}(\eta_i) = v(\mathbf{X}_i)(1 + \Delta_i)$, see Ruppert et al. (1997, p. 263). This motivates us to use the standardized residuals $\tilde{\eta}_i = \eta_i/(\hat{v}(\mathbf{X}_i)(1 + \Delta_i))^{1/2}$.

By the role the covariance-function plays in the calculation of e.g. the variance of a linear combination of random variables, it is necessary that it possess the property of positive definiteness. A number of different techniques is available for estimation of proper covariance-functions. However, for simplicity, it is recommended to fit a parametric function that is known to be positive definite. Other techniques include fitting a sum of Bessel functions, see Shapiro and Botha (1991), or using nonparametric estimation with subsequent transformation into a valid function, see e.g. Christakos (1984) and Hall et al. (1994).

Finally, by (1) the estimate of the covariance matrix, $\boldsymbol{\Sigma}$, is constructed by,

$$\widehat{\boldsymbol{\Sigma}}_{ij} = \hat{\rho}(||\mathbf{X}_i - \mathbf{X}_j||)\hat{v}(\mathbf{X}_i)^{1/2}\hat{v}(\mathbf{X}_j)^{1/2}. \tag{24}$$

2.4.6 Algorithm for Bandwidth Selection

The following is a review of the algorithm for selecting bandwidth matrices using EBBS_{dep}. The Software, written for MATLAB 6, is available at the web address: `http://www.maths.lth.se/matstat/staff/torgny`.

For each location $\mathbf{x}_l$ on a subspace of G_x, denoted G'_x, of size $g_1 \times g_2$ the bandwidth matrix, $\mathbf{H}(\mathbf{x}_l)$, corresponding to the least estimated MSE among the matrices with triples in $G'_U(\mathbf{x}_l) \subset G_U(\mathbf{x}_l)$ is chosen. Here $G_U(\mathbf{x}_l)$ is supplied by the user by specifying a lower and upper bound, q_1 and q_2, for the spans of h_1 and h_2 and by specifying the size of the grid of bandwidths to try, $M_1 = (M_{11}, M_{12}, M_{13})$. In this way $G_U(\mathbf{x}_l) = \{(h_{1,j_1}(\mathbf{x}_l), h_{2,j_2}(\mathbf{x}_l), h_{3,j_3}^{j_1,j_2}(\mathbf{x}_l))\}, j_1 = 1, \ldots, M_{11}; j_2 = 1, \ldots, M_{12}; j_3 = 1, \ldots, M_{13}$. Remember that in order to ensure positive definite bandwidth matrices the ranges of valid values for $h_{12}(\mathbf{x}_l)$ are dependent on the corresponding values of $h_1(\mathbf{x}_l)$ and $h_2(\mathbf{x}_l)$, hence the superscript on $h_{3,j_3}^{j_1,j_2}(\mathbf{x}_l)$. The set $G'_U(\mathbf{x}_l)$ is defined below.

For a location $\mathbf{x}_l \in G'_x$ and a bandwidth matrix $\mathbf{H}_0(\mathbf{x}_l)$, where $\mathbf{H}_0(\mathbf{x}_l)$ consists of $(h_{1,j_0}(\mathbf{x}_l), h_{2,j_0}(\mathbf{x}_l), h_{12,j_0}(\mathbf{x}_l)) \in G'_U(\mathbf{x}_l)$, the estimate of the MSE of $\hat{m}_{k_1,k_2}(\mathbf{x}_l; \mathbf{H}_0, p)$ is calculated by using (15) for the bias and (16) for the variance,

$$\begin{aligned}\widehat{\mathsf{MSE}}(\mathbf{x}_l; \mathbf{H}_0) = \widehat{\mathsf{Var}}\{\hat{m}_{k_1,k_2}(\mathbf{x}_l; \mathbf{H}_0, p)\} \\ + [\hat{c}_{p-K+1}(\mathbf{x}_l) \operatorname{tr}(\mathbf{H}_0^{\mathsf{T}}\mathbf{H}_0)^{(p-K+1)/2} + \ldots \\ + \hat{c}_{p-K+t}(\mathbf{x}_l) \operatorname{tr}(\mathbf{H}_0^{\mathsf{T}}\mathbf{H}_0)^{(p-K+t)/2}]^2.\end{aligned} \tag{25}$$

Here $\widehat{\mathsf{Var}}\{\hat{m}_{k_1,k_2}(\mathbf{x}_l; \mathbf{H}_0, p)\}$ is given by plugging $\widehat{\boldsymbol{\Sigma}}$ into (16).

When fitting the polynomial defined in (14) for bias estimation at $\mathbf{x}_l$ and for $\mathbf{H}_0(\mathbf{x}_l)$ the following triples are used, $G_{H_0}(\mathbf{x}_l) = \{(h_{1,j_1}(\mathbf{x}_l), h_{2,j_2}(\mathbf{x}_l), h_{3,j_3}^{j_1,j_2}(\mathbf{x}_l))\}, j_1 = j_0 - J_1, \ldots, M_{11} - J_2; j_2 = j_0 - J_1, \ldots, M_{12} - J_2; j_3 = j_0 - J_1, \ldots, M_{13} - J_2$, where J_1 and J_2 are integers such that $J_1 + J_2 \geq t$ and $J_2 \geq 0$. Thus bias estimation is performed for bandwidth matrices with triples in $G'_U(\mathbf{x}_l) = \{(h_{1,j_1}(\mathbf{x}_l), h_{2,j_2}(\mathbf{x}_l), h_{3,j_3}^{j_1,j_2}(\mathbf{x}_l))\}, j_1 = 1 + J_1^*, \ldots, M_{11} - J_2; j_2 = 1 + J_1^*, \ldots, M_{12} - J_2; j_3 = 1 + J_1^*, \ldots, M_{13} - J_2$, where $J_1^* = \max\{0, J_1\}$.

This method for bias estimation is rather slow since it is involves $(J_1 + J_2 + 1)^3$ mean estimates for each $\mathbf{H}_0(\mathbf{x}_l)$. However as seen above the elements in $G_{H_0}(\mathbf{x}_l)$ are chosen from $G_U(\mathbf{x}_l)$. In this way there is a large intersection between $G_{H_i}(\mathbf{x}_l)$ and $G_{H_j}(\mathbf{x}_l)$ for neighbouring $\mathbf{H}_i(\mathbf{x}_l)$ and $\mathbf{H}_j(\mathbf{x}_l)$ so that several mean function estimates can be reused in subsequent bias estimates.

Next the estimates of the mean squared error of $\hat{m}_{k_1,k_2}(\mathbf{x}_l; \mathbf{H}_0, p)$ for $\mathbf{H}_0(\mathbf{x}_l)$ with triples in $G'_U(\mathbf{x}_l)$ are interpolated by linear interpolation to a finer grid, $G^*_U(\mathbf{x}_l) = \{(h^*_{1,j_1}(\mathbf{x}_l), h^*_{2,j_2}(\mathbf{x}_l), h_{3,j_3}^{*,j_1,j_2}(\mathbf{x}_l))\}, j_1 = 1, \ldots, M_{21}; j_2 = 1, \ldots, M_{22}; j_3 = 1, \ldots, M_{23}$, of size $M_2 = (M_{21}, M_{22}, M_{23})$ where in order to avoid extrapolation $h^*_{k,1}(\mathbf{x}_l) \geq h_{k,1+J_1^*}(\mathbf{x}_l)$ and $h^*_{k,M_{2k}}(\mathbf{x}_l) \leq h_{k,M_{1k}-J_2}(\mathbf{x}_l)$, $k = 1, 2$.

Depending on the sizes of $G'_U(\mathbf{x}_l)$ and $G^*_U(\mathbf{x}_l)$ the interpolated estimated MSE can be quite rough and may therefore be smoothed. The smoothed $\widehat{\mathsf{MSE}}(\mathbf{x}_l; \mathbf{H}_0)$ is a local weighted average using trivariate spherically contoured Epanechnikov weights. The proposed bandwidth matrix at location $\mathbf{x}_l$ is then the matrix $\widehat{\mathbf{H}}(\mathbf{x}_l)$ that minimizes the (possibly smoothed) $\widehat{\mathsf{MSE}}(\mathbf{x}_l; \mathbf{H}_0)$.

According to Ruppert (1997) (considering selection of scalar bandwidths), it is important that the first local minimum of the estimated MSE is used. The argument for this is that this method for bias estimation underestimates bias when the bandwidths are so large that features of $\hat{m}_{k_1,k_2}(\cdot;\mathbf{H},p)$ are smoothed away. Thus the estimated bias tends to zero as the bandwidth tends to infinity. In EBBS_{dep} the local minimum that is closest to the origin is used. Of course it may happen that two or more local minima are located at the same distance from the origin. In this case the bandwidths of the the first local minimum are selected. Unfortunately this method for bandwidth selection overcompensates for the underestimation of the bias so that the diagonal components of the selected $\mathbf{H}$ are too small. However for reasonably short ranged dependence this selection method seems superior to using the global minimum, as can be seen in Section 3. In the algorithm $\mathbf{H}_{\min}(\mathbf{x}_l)$ constitutes a local minimum if the value of $\widehat{\mathsf{MSE}}(\mathbf{x}_l;\mathbf{H}_{\min})$ is less than any other value of the estimated MSE in the $(2\,kopt+1)\times(2\,kopt+1)$ region around $\mathbf{H}_{\min}(\mathbf{x}_l)$.

In Section 3 we study the performance of this method for bandwidth selection in a simulation experiment. There is also a comparison with bandwidths based on the global minimum of the estimated MSE, which will be denoted by $\text{EBBS}_{\text{dep}}^{\text{GM}}$.

Even though the estimated MSE might be smoothed, the components of the bandwidth matrices $\mathbf{H}_{\min}(\mathbf{x})$ are quite rough as functions of the location. Therefore they are smoothed by local weighted averages using bivariate spherically contoured Epanechnikov weights. Finally the bandwidth components are interpolated from G'_x into G_x by cubic interpolation.

3 Simulations

Studies were performed for evaluation of the dependence adjusted bandwidth selector, EBBS_{dep}, on simulated data displaying varying degrees of dependence. The purpose of these experiments was to observe the performance of the bias estimation procedure and its effects on the bandwidth selection procedure.

The following are common to the studies. Kernels are based on the spherically contoured Epanechnikov kernel,

$$K(\mathbf{x}) = \max\{0, \frac{2}{\pi}(1-||\mathbf{x}||^2)\}, \tag{26}$$

which is optimal in the sense that it minimizes the asymptotically MSE, see Fan et al. (1997, Theorem 3.1, p. 88). All bandwidths including the optimal ones were calculated on the same grid, G_x, a regular grid of size 20×20 contained in $[0,1]\times[0,1]$. At location $\mathbf{x}'_i \in G_x$ the optimal bandwidth matrix, $\mathbf{H}_{\text{opt}}(\mathbf{x}'_i)$, minimizes the

MSE of the $\hat{m}_{k_1,k_2}(\mathbf{x}'_i; \mathbf{H}, p)$ estimate,

$$\begin{aligned}\mathsf{MSE}\{\mathbf{x}'_i, \mathbf{H}\} = e_i^{\mathsf{T}}[&(\mathbf{S}_{G_x}(l)\mathbf{m} - \mathbf{m}_{G_x}^{(k_1,k_2)}) \odot (\mathbf{S}_{G_x}(l)\mathbf{m} - \mathbf{m}_{G_x}^{(k_1,k_2)}) \\ &+ D_K^2 \ \mathrm{Diag}\{\mathbf{S}_{G_x}(l)\mathbf{\Sigma}\mathbf{S}_{G_x}^{\mathsf{T}}(l)\}],\end{aligned} \tag{27}$$

where $\mathbf{m}_{G_x}^{(k_1,k_2)}$ is the column vector containing the (k_1, k_2):th derivative of $m(\cdot)$ at the locations in G_x and $\mathbf{m} = (m(\mathbf{X}_1), \ldots, m(\mathbf{X}_n))^{\mathsf{T}}$. For each $i = 1, \ldots, m$, the $\mathsf{MSE}\{\mathbf{x}'_i, \mathbf{H}\}$ is minimized over a grid of 30×30 (h_1, h_2) pairs for the diagonal $\mathbf{H}$ and minimized over a grid of $30 \times 30 \times 29$ (h_1, h_2, h_{12}) triples for the full bandwidths matrices. Here each of h_1 and h_2 are equally spaced on the logarithmic scale between bandwidths that correspond to spans 0.15 and 1. For h_{12} the values are symmetric about zero and the absolute of these values equally spaced on the log-scale. The reason for having an odd number of grid-points for h_{12} is that this grid is symmetric about 0 and that we want $h_{12} = 0$ to be included. Since the resulting optimal bandwidths are quite rough they are smoothed in the same way as the bandwidths in the simulations.

To measure the performance of the different bandwidth selectors the mean average squared error (MASE) is used. The MASE using bandwidth matrices $\mathbf{H}(\mathbf{x}'_i)$, $i = 1, \ldots, m$ for estimation at the locations in G_x is given by,

$$\begin{aligned}\mathsf{MASE}(\{\mathbf{H}(\mathbf{x}'_i)\}_1^m) = &\frac{1}{m}(\mathbf{S}_{G_x}(l)\mathbf{m} - \mathbf{m}_{G_x}^{(k_1,k_2)})^{\mathsf{T}}(\mathbf{S}_{G_x}(l)\mathbf{m} - \mathbf{m}_{G_x}^{(k_1,k_2)}) \\ &+ \frac{1}{m}\ \mathrm{tr}\{\mathbf{S}_{G_x}(l)\mathbf{\Sigma}\mathbf{S}_{G_x}^{\mathsf{T}}(l)\}.\end{aligned} \tag{28}$$

3.1 Empirical-Bias Estimation and Bandwidth Selection Performance

In this simulation study the performance of the empirical-bias estimator and the bandwidth selection method is evaluated. For this purpose, normal random fields of size $n = 900$ were generated. Following (1) we used the mean-function, $m(\mathbf{Ax})$, which is a clockwise rotation by $\vartheta = \pi/6$ radians,

$$\mathbf{A} = \begin{pmatrix} \cos\vartheta & -\sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{pmatrix}, \tag{29}$$

of the additive mean-function,

$$m(\mathbf{x}) = \sin(2\pi x_1) + 4(x_2 - 0.5)^2. \tag{30}$$

The variance-function was constant, $v(\cdot) = 0.8$. The $\mathbf{X}_i$'s were iid uniform random variables on the unit square and the dependence was modeled by the exponential correlation function,

$$\rho(\mathbf{X}_i - \mathbf{X}_j) = \exp(-a\,||\mathbf{X}_i - \mathbf{X}_j||), \ \text{ with } a > 0. \tag{31}$$

Estimation was performed on the regularly spaced 20×20 grid, G_x, defined above.

In order to study the effects of the dependence in the model, different values of the parameter a were used, $a = 5$ (strong correlation), $a = 20$ (relatively strong correlation), $a = 40$ (relatively weak correlation) and $a = 200$ (approximately uncorrelated). For each value of a fifty random fields were simulated.

In this experiment we study bandwidth selection for the local linear estimate of the mean function, $\hat{m}_{0,0}(\cdot; \mathbf{H}, 1)$. Comparisons were made for the EBBS, EBBS_{dep}, $\text{EBBS}_{\text{dep}}^{\text{GM}}$ and $\mathbf{H}_{\text{opt}}$ bandwidths using both diagonal bandwidth matrices with h_1 and h_2 on the diagonal and full bandwidth matrices. In the following, procedures for selection of full bandwidth matrices will be superscripted by full, e.g. $\text{EBBS}_{\text{dep}}^{\text{full}}$. Also, the procedure termed EBBS here is based on the software for EBBS_{dep} using the diagonalized $\mathbf{\Sigma}$.

For EBBS, EBBS_{dep} and $\text{EBBS}_{\text{dep}}^{\text{GM}}$ the following parameters were used, $(g_1, g_2) = (15, 15)$, $M1 = (20, 20)$, $M2 = (35, 35)$, $J_1 = 1$, $J_2 = 2$, $t = 2$, $\mathit{msespan} = (0, 0)$, i.e. no smoothing of the estimated MSE fields and $\mathit{bandspan} = (0.15, 0.15)$ specifying a span of 0.15 in both directions for smoothing the bandwidths. The same parameters were used for $\text{EBBS}^{\text{full}}$, $\text{EBBS}_{\text{dep}}^{\text{full}}$ and $\text{EBBS}_{\text{dep}}^{\text{full,GM}}$ except for the parameter M_1 which was equal to $(20, 20, 15)$. For EBBS, EBBS_{dep}, $\text{EBBS}^{\text{full}}$ and $\text{EBBS}_{\text{dep}}^{\text{full}}$ with the values of the parameters given above, the values $1, \ldots, 4$ for kopt were tested. Based on the MASE measures of these tests $\mathit{kopt} = 3$ was chosen for these experiments. For each type of correlation the means of the selected bandwidths, the mean function estimates and the MASE's were collected.

Figure 1 shows in (i) a typical realization of a normal random field based on the exponential correlation-function with parameter $a = 5$. The mean-function used for these simulations are shown in Figure 1(ii). Also in Figure 1 are shown examples of $\hat{m}_{0,0}(\cdot; \mathbf{H}, 1)$, for bandwidths based on $\text{EBBS}_{\text{dep}}^{\text{full,GM}}$ in (iii) and $\text{EBBS}^{\text{full}}$ in (iv). Here the bandwidths selected by EBBS are too small which results in a rough estimate of the mean-function while the bandwidths based on $\text{EBBS}_{\text{dep}}^{\text{full,GM}}$ gives a more smooth and accurate estimate.

For $a = 5$ the means of h_1 for $\text{EBBS}^{\text{full}}$, $\text{EBBS}_{\text{dep}}^{\text{full}}$ and $\text{EBBS}_{\text{dep}}^{\text{full,GM}}$ are shown in Figure 2 together with the optimal values of h_1. Corresponding results for h_2 are given in Figure 3. Clearly the bandwidths for $\text{EBBS}^{\text{full}}$ and $\text{EBBS}_{\text{dep}}^{\text{full}}$ are generally too small. As expected the simulations showed that EBBS and $\text{EBBS}^{\text{full}}$ are roughly unaffected by the level of dependence. $\text{EBBS}_{\text{dep}}^{\text{full,GM}}$, on the other hand seem to capture more of the structure of the optimal bandwidths but lack the dynamics of the latter, resulting in bandwidths that are too large on average. This is an effect of the underestimation of the bias, as was noted in Section 2.4.

Consistently over the different values of a the bandwidths selected by EBBS_{dep} and $\text{EBBS}_{\text{dep}}^{\text{full}}$ are generally too small. As discussed above this is caused by using the "first" local minimum of the estimated MSE.

Concerning the off-diagonal elements for $a = 5$ which are presented in Figure 4, $\text{EBBS}^{\text{full}}$ and $\text{EBBS}_{\text{dep}}^{\text{full}}$ fail to find the structure of the optimal h_{12}. On the other hand, h_{12} selected by $\text{EBBS}_{\text{dep}}^{\text{full,GM}}$ are much closer to the optimal here. This behaviour was consistent for all four values of a.

Tables 1 and 2 summarizes the simulation study by use of the means of the MASE measures. In Table 1 the means of the MASE's corresponding to the different levels of dependence for EBBS, EBBS_{dep} and $\text{EBBS}_{\text{dep}}^{\text{GM}}$ using diagonal bandwidth matrices as well as the MASE for the smoothed optimal bandwidths are presented. Corresponding values for full bandwidth matrices are given in Table 2.

These results show that EBBS_{dep} and $\text{EBBS}_{\text{dep}}^{\text{full}}$ are closer to the optimal bandwidths than $\text{EBBS}_{\text{dep}}^{\text{GM}}$ and $\text{EBBS}_{\text{dep}}^{\text{full,GM}}$ for short range correlations. Thus EBBS_{dep} and $\text{EBBS}_{\text{dep}}^{\text{full}}$ are recommended for heteroscedastic models where $v(\cdot)$ needs to be estimated separately. The reason is that the dependence is generally of short range in such models.

For illustration of the performance of the empirical-bias estimator, means of the estimated bias by $\text{EBBS}_{\text{dep}}^{\text{full}}$ and $\text{EBBS}_{\text{dep}}^{\text{full,GM}}$ in the simulations for $a = 5$ are presented in Figure 5. This figure shows the means of the estimated bias as well as the means of the true bias. It shows that empirical-bias estimation underestimates the level of the bias though it is most severely underestimated by $\text{EBBS}_{\text{dep}}^{\text{full,GM}}$.

In Figure 6 the simulation means of the MSE for the different methods are shown for $a = 5$. From this figure it seems that $\text{EBBS}_{\text{dep}}^{\text{full,GM}}$ are more affected by boundary effects than the other methods. Note that this boundary issue has an influence also on the values of the MASE's in Tables 1 and 2.

In order to assess the performance of EBBS_{dep} in a situation where the dependence between the errors are estimated, another simulation experiment was conducted where $\mathbf{\Sigma}$ was estimated for each of the four cases of a. In this experiment $v(\cdot)$ was assumed to be constant and equal to one so that no estimates of this function were necessary. The covariance-function in the model was estimated by parametric variogram fitting to the empirical semi-variogram,

$$\hat{\gamma}(\tau) = \frac{1}{2n(\tau, t)} \sum_{(i,j) \in S(\tau,t)} (\eta_i - \eta_{i'})^2, \tag{32}$$

where $S(\tau, t) = \{(i, i') : \tau - t \leq ||\mathbf{X}_i - \mathbf{X}_{i'}|| < \tau + t\}$ is the set of indices for which the locations are at a distance $\tau \pm t$ of each other and $n(\tau, t)$ is the number of elements in $S(\tau, t)$. The residuals, η_i, were calculated by making an initial smooth of $\mathbf{Y}$ using local linear regression with a constant bandwidth matrix,

$$\mathbf{H}_{\text{init}} = 1.7 \begin{pmatrix} \hat{\sigma}_{X_1} & 0 \\ 0 & \hat{\sigma}_{X_2} \end{pmatrix}, \tag{33}$$

where $\hat{\sigma}_{X_1}$ and $\hat{\sigma}_{X_2}$ are the estimated standard deviations of the respective components of the $\mathbf{X}_i$'s.

How well EBBS$_{\text{dep}}$ performs in this case depends on whether the form of the chosen parametric $\rho(\cdot)$ is appropriate and how well the parameters of $\rho(\cdot)$ are estimated. In this experiment the exponential variogram function is used for estimation in order to study the latter aspect. For calculation of the experimental semi-variogram the $\{(\eta_i, \eta_{i'}), i \neq i'\}$ pairs were partitioned into 80 intervals of equal length. The first 50 of these were used for variogram fitting. The parameters a and b were estimated following the weighted least-squares minimization of Cressie (1993, p. 96). Table 3 shows the means of the parameter estimates as a function of a. Recall that the true value of b is 0.8.

From this table it can be concluded that the components of the chosen $\mathbf{H}_{\text{init}}$ are too small to accurately estimate the parameters for the case $a = 5$. For the other values of a $\mathbf{H}_{\text{init}}$ is more suitable though the correlation is overestimated for $a = 40$ and $a = 200$.

Table 4 displays $\mathbf{H}_{\text{opt}}$ for every value of a together with the averages of the MASE's for EBBS, EBBS$_{\text{dep}}$ and EBBS$_{\text{dep}}^{\text{GM}}$ based on the estimated $\boldsymbol{\Sigma}$'s. In Table 5 the corresponding values for the mean-function estimates based on full bandwidth matrices are given. Except for EBBS$_{\text{dep}}^{\text{full,GM}}$ the average MASE's using estimated parameter values for $\rho(\cdot)$ are close to the corresponding values when using the true $\boldsymbol{\Sigma}$.

3.2 Variance-function estimation

A small simulation study was performed to assess the effects of the choice of bandwidths for estimation of $m(\cdot)$ on the variance-function estimate. A similar study for the univariate variance-function estimator based on EBBS is presented in Ruppert et al. (1997). The conclusion there is that it is quite insensitive as long as the bandwidths for the initial smooth are not so large such that the bias of $\hat{m}(\cdot)$ inflates the residuals causing positive bias.

In this study iid errors and a constant variance-function, $v(\cdot) = 1$, are used since we are interested in the error due to the initial smooth, not the error due to a non-constant $v(\cdot)$ or due to correlation between the ε_i's. A simulation study of covariance-estimation as functions of location with fixed displacement τ, including the variance-function estimator for $\tau = 0$, conducted to observe the influence of the initial smooth is presented in Lindström (2003). In that study (using stationary data) it is apparent that the initial smooth is the major source for bias.

The set up of the simulation experiment was as follows: 50 samples of size $n = 900$ were generated using the mean-function,

$$m(\mathbf{x}) = 10 \exp\{-20[(x_1 - 0.5)^2 + (x_2 - 0.5)^2]\}. \tag{34}$$

The errors were iid $N(0, 1)$ samples and the $\mathbf{X}_i$'s were iid uniform random variables on the unit square. Estimation was performed using diagonal bandwidth matrices and there were four different choices of bandwidths for the initial smooth, spans (0.15, 0.15), (0.4, 0.4) and (0.75, 0.75) as well as EBBS_{dep} bandwidths. Here the spans for EBBS_{dep} were ranging from span (0.15, 0.15) to span (1, 1) and the span for estimation of $m(\cdot)$ in step 1 were (0.15, 0.15). For these data sets the iteration sequence in the estimation procedure was not found to be necessary.

For each sample and for each of the four different choices of bandwidths for the initial smooth the mean-function was estimated on the locations for the observations by using local polynomials of order $p_1 = 2$. The subsequent smoothing of the squared residuals were performed on a 20 by 20 regularly spaced grid, G_x, on the unit square and was based on local polynomials of order $p_2 = 1$ and EBBS_{dep} bandwidths for spans in the range (0.15, 0.15) to (1, 1). Figure 7 shows a typical realization together with the mean-function and its estimates based on bandwidths of spans (0.15, 0.15) and (0.4, 0.4).

Bias and standard deviations for each of the four $\hat{v}(\cdot)$'s were calculated based on the 50 simulations. These are presented in Figures 8 and 9, respectively. From these figures it can be seen that the bias are quite low so that the standard deviations are dominating the errors except for the case with large bandwidths for the initial smooth. In this case the bias of the $m(\cdot)$-estimate is so large that it has inflated the bias of $\hat{v}(\cdot)$. Regarding the standard deviations for the other cases we see that they are about the same except at the boundary for the case with extremely low initial bandwidths.

Table 6 below contains the means of the MASE's for the different choices of bandwidths for the initial mean estimate. From this table and Figures 8 and 9 it seems that EBBS_{dep} is a good candidate for bandwidth selection for the initial smooth.

4 Summary

We have presented a method for data-driven selection of local bandwidth matrices for local polynomial regression. The method is based on the empirical-bias bandwidth selector of Ruppert (1997), which is constructed for scalar bandwidths and independent data. This paper presents an extension of EBBS allowing for full bandwidth matrices and correlated data.

Bandwidths are chosen by minimization of an estimate of the MSE, which is constructed from separate estimates of the variance and the squared bias of the estimator, $\hat{m}_{k_1,k_2}(\mathbf{x};\mathbf{H},p)$. For the calculation of the variance estimate the known expression from weighted least squares theory is used. The only parameter in this expression that needs to be estimated is the covariance matrix. This estimate is constructed from separate estimates of the variance-function and the correlation-function of the errors.

Bias estimation is performed empirically where the form of the estimator is based on asymptotic expressions for the bias of $\hat{m}_{k_1,k_2}(\mathbf{x};\mathbf{H},p)$. Simulations show that bias estimation is a sensitive part of the EBBS_{dep} procedure. A problem with the empirical bias-estimation is that it tends to zero as $\text{tr}(\mathbf{H}) \to \infty$. Consequently this needs to be corrected for or the selected bandwidths will be too large. In this paper, in order to compensate for this problem, selected bandwidths are those of the local minimum of the estimated MSE closest to the origin. This is the major drawback of this method and is a topic for possible future research.

A procedure for estimating the variance-function in the presence of short ranged correlated errors was presented. Based on the local polynomial variance-function estimator proposed by Ruppert et al. (1997) adjustments are made to allow for EBBS_{dep} controlled bandwidths for smoothing of the squared residuals. Results for the bias and variance of this estimator are also presented.

The form of the bias estimate is motivated by the asymptotic results of Ruppert and Wand (1994). However, they provided no results for the case of derivative estimation in the multivariate case. The conditional bias of the estimate of the partial derivatives of order one using local quadratic polynomials are given in the Appendix.

Two simulation experiments have been performed for assessment of the estimation procedure of this paper. In these experiments separate foci have been on bias estimation and the subsequent bandwidth selector in the first one and on the influence of the initial smooth on the performance of the variance-function estimator in the second one. In this way the performance of the different estimates in the procedure could be more easily evaluated.

The first simulation experiment confirmed that the squared bias estimate is negatively biased, which is reflected in the selected bandwidths. However, comparisons with bandwidths selected from the local minimum of the estimated MSE showed that in the case of strong correlation this might be a better choice. For short ranged correlations bandwidths based on the the "first" local minima are better.

The second simulation experiment studied the influence of the selection of bandwidths for the initial estimate of $m(\cdot)$ on the performance of the local polynomial variance-function estimator. In this experiment the variance-function estimator performed quite well as long as the bandwidths of the initial smooth were reasonably small. The procedure for using EBBS_{dep} also for the initial bandwidths performed well here and can be recommended.

The EBBS_{dep} software is written for MATLAB 6 and is available at the web address: `http://www.maths.lth.se/matstat/staff/torgny`. Estimation of each of the mean-functions in Section 3.1 using the known $\mathbf{\Sigma}$ takes for EBBS_{dep} about four minutes to run on a Pentium IV 2600 PC and about 85 minutes for $\text{EBBS}_{\text{dep}}^{\text{full}}$. Depending on the data and the shape of the region for estimation the stability of the software can be quite sensitive to the values of the parameters. Often some effort must be put into tuning the parameters for best result, though in our experience the procedure works well when well tuned.

Appendix

The following result is taken from Lindström (2003), in that paper spatial covariance-function estimation is performed using smoothed residual products under the assumption of local second-order stationarity and isotropy. That procedure is aimed at estimation of covariances as functions of the location while the displacement τ is fixed. Here the special case of variance-function estimation is treated. Furthermore, the results are slightly modified to allow for estimation on a general grid, G_x, containing m locations. The changes in the proofs are straightforward and will thus not be presented.

Theorem 1 *Let* $\mathbf{b}_1 = (\mathbf{S}_1 - \mathbf{I})\mathbf{m}$ *denote the bias vector of the initial smooth and assume that the error field* $\{\varepsilon_i\}_{i=1}^n$ *is a stationary isotropic random field. Then*

$$\begin{aligned} \mathsf{E}\{\hat{\mathbf{v}} - \mathbf{v}_{G_x} | \mathbf{X}_1, \ldots, \mathbf{X}_n\} = & \frac{\mathbf{S}_2\mathbf{v} - (\mathbf{1} + \mathbf{S}_2\boldsymbol{\Delta}) \odot \mathbf{v}_{G_x}}{\mathbf{1} + \mathbf{S}_2\boldsymbol{\Delta}} \\ & + \frac{\mathbf{S}_2[\mathbf{b}_1 \odot \mathbf{b}_1 + \mathrm{Diag}\{\mathbf{S}_1\boldsymbol{\Sigma}\mathbf{S}_1^\mathsf{T} - \mathbf{S}_1\boldsymbol{\Sigma} - \boldsymbol{\Sigma}\mathbf{S}_1^\mathsf{T}\}]}{\mathbf{1} + \mathbf{S}_2\boldsymbol{\Delta}}, \end{aligned} \tag{35}$$

where $\mathbf{v} = (v(\mathbf{X}_1), \ldots, v(\mathbf{X}_n))^\mathsf{T}$, $\mathbf{v}_{G_x}$ *is the column vector containing the values of* $v(\cdot)$ *evaluated at the locations in* G_x, $\mathbf{1}$ *is a* $m \times 1$ *vector of ones and* $\boldsymbol{\Delta} = \mathrm{Diag}\{\mathbf{S}_1\mathbf{S}_1^\mathsf{T} - 2\mathbf{S}_1\}$.

If the error field follows a Gaussian distribution,

$$\begin{aligned} \mathsf{Cov}(\hat{\mathbf{v}} | \mathbf{X}_1, \ldots, \mathbf{X}_n) = & \mathbf{S}_2[2\,\{(\mathbf{S}_1 - \mathbf{I})\boldsymbol{\Sigma}(\mathbf{S}_1 - \mathbf{I})^\mathsf{T}\} \odot \{(\mathbf{S}_1 - \mathbf{I})\boldsymbol{\Sigma}(\mathbf{S}_1 - \mathbf{I})^\mathsf{T}\} \\ & + 4\,(\mathbf{b}_1\mathbf{b}_1^\mathsf{T}) \odot \{(\mathbf{S}_1 - \mathbf{I})\boldsymbol{\Sigma}(\mathbf{S}_1 - \mathbf{I})^\mathsf{T}\}]\mathbf{S}_2^\mathsf{T} \\ & /(\{\mathbf{1} + \mathbf{S}_2\boldsymbol{\Delta}\}\{\mathbf{1} + \mathbf{S}_2\boldsymbol{\Delta}\}^\mathsf{T}). \end{aligned} \tag{36}$$

$\square$

Asymptotic results for the conditional bias of the first order partial derivative estimates using multivariate local quadratic regression are derived below. These are given for covariates of a general dimension, d. That is, assume a random sample $\{(\mathbf{X}_i, Y_i),\ i = 1, \ldots, n\}$ satisfying model (1) with the exception that for each response variable Y_i the location is described by the predictor variable, $\mathbf{X}_i$, having bounded density $f(\cdot)$ with support $\mathrm{supp}\{f(\cdot)\} \subseteq \mathbb{R}^d$. Also, let $m^{(k)}(\cdot)$ and $\hat{m}_k(\cdot; \mathbf{H}, p)$, where $\mathbf{k} = (k_1, \ldots, k_d)$, expand the notation of $m^{(k_1,k_2)}(\cdot)$ respectively $\hat{m}_{k_1,k_2}(\cdot; \mathbf{H}, p)$ to d dimensions in the natural way. Note that $\mathbf{k}$ is not printed in bold when used in sub and superscripts.

Let us begin by introducing some necessary conditions which are taken from Ruppert and Wand (1994). Note that their condition (A2) is not used here.

(A1) The kernel $K(\cdot)$ is a compactly supported, bounded kernel with second-order moments $\int \mathbf{u}\mathbf{u}^\mathsf{T} K(\mathbf{u}) d\mathbf{u} = \mu_2(K)\mathbf{I}_d$, where $\mu_2(K) \neq 0$ is scalar and $\mathbf{I}_d$ is the

$d \times d$ identity matrix. In addition, all odd-order moments of $K(\cdot)$ vanish, that is, $\int u_1^{l_1} \cdots u_d^{l_d} K(\mathbf{u}) d\mathbf{u} = 0$ for all nonnegative integers $l_1, \ldots, l_d$ such that their sum is odd. The last condition is satisfied by spherically symmetric kernels and products of univariate symmetric kernels.

(A3) The bandwidth matrices depend on n and the sequence of $\mathbf{H}$ is such that $n^{-1}|\mathbf{H}|^{-1}$ and each entry of $\mathbf{H}$ tends to zero as $n \to \infty$ with $\mathbf{H}$ remaining symmetric and positive definite. Also, there is a fixed constant L such that $\lambda_{\max}(\mathbf{H})/\lambda_{\min}(\mathbf{H}) < L$ for all n, where $\lambda_{\max}(\mathbf{H})$ and $\lambda_{\min}(\mathbf{H})$ are the largest and the smallest eigenvalue of $\mathbf{H}$ respectively.

(A4) There is a convex set $\mathcal{C}$ with nonnull interior and containing a point $\mathbf{x}_\partial$ on the boundary of the support of $f(\cdot)$ such that

$$\inf_{\mathbf{x} \in \mathcal{C}} f(\mathbf{x}) > 0. \tag{37}$$

As mentioned previously, kernel functions are often taken to be d-variate density functions so that $\int \mathbf{u} K(\mathbf{u}) d\mathbf{u} = 0$. Then according to assumption (A1) the covariance matrix of $K(\cdot)$ is $\mu_2(K)\mathbf{I}_d$. The convergence rate of the bandwidths in assumption (A3) ensures consistent estimates. The eigenvalues of $\mathbf{H}$ are the lengths of the axis of the ellipsoid volume that defines the local neighbourhood centered at $\mathbf{x}$. The requirement that $\lambda_{\max}(\mathbf{H})/\lambda_{\min}(\mathbf{H}) < L$ ensures that the entries of $\mathbf{H}$ tend to zero at the same rate so that the local neighbourhood does not collapse into an elliptic plane as n tends to infinity. Also, since the $\mathbf{X}_i$'s are random the bandwidth matrices should depend on $\mathbf{X}_1, \ldots, \mathbf{X}_n$. Otherwise there is a positive probability that none of the $\mathbf{X}_i$'s are in the local neighbourhood of $\mathbf{x}$ so that estimation of $m^{(k)}(\mathbf{x})$ is not possible to perform. However, this dependence will not be explicitly expressed. Assumption (A4) ensures non-degenerate integrals at boundary points, $\mathbf{x}_\partial$.

We need the following definitions. For local quadratic regression the design matrix is given by,

$$\mathbf{X}_{2,x} = \begin{pmatrix} 1 & (\mathbf{X}_1 - \mathbf{x})^\mathsf{T} & \mathrm{vech}^\mathsf{T}\{(\mathbf{X}_1 - \mathbf{x})(\mathbf{X}_1 - \mathbf{x})^\mathsf{T}\} \\ \vdots & \vdots & \vdots \\ 1 & (\mathbf{X}_n - \mathbf{x})^\mathsf{T} & \mathrm{vech}^\mathsf{T}\{(\mathbf{X}_n - \mathbf{x})(\mathbf{X}_n - \mathbf{x})^\mathsf{T}\} \end{pmatrix}, \tag{38}$$

where $\mathrm{vech}(\mathbf{A})$ is the $d(d+1)/2 \times 1$ vector of elements in $\mathrm{vec}(\mathbf{A})$ where all the elements above the diagonal in $\mathbf{A}$ has been removed and $\mathrm{vec}(\mathbf{A})$ is the column vector of entries in the matrix $\mathbf{A}$ taken columnwise.

Let $\nabla_g(\mathbf{x})$ denote the $d \times 1$ vector of the first-order partial derivatives of the d-variate (sufficiently smooth) function $g(\cdot)$ evaluated at $\mathbf{x}$ and $\mathcal{H}_g(\mathbf{x})$ denote the $d \times d$ Hessian matrix of second-order partial derivatives of $g(\cdot)$. Also, for any real-valued function $g(\cdot)$ of $\mathbf{x}$ the $\mathbf{k}$:th-order differential at $\mathbf{x}_0$ is,

$$(d^k_{x_0})g(\mathbf{x}) = \sum_{k_1+\ldots+k_d=K} \binom{K}{k_1 \ldots k_d} x_1^{k_1} \ldots x_d^{k_d} \left. \frac{\partial g(\mathbf{x})}{\partial x_1^{k_1} \ldots \partial x_d^{k_d}} \right|_{\mathbf{x}=\mathbf{x}_0}, \tag{39}$$

if all the Kth-order partial derivatives exist.

Now, consider estimation of the first order partial derivative in direction l, $l = 1, \ldots, d$. Let $\mathbf{k}_l$ be a $1 \times d$ vector of zeros except for position l which equals one. Then,

$$\mathsf{E}\{\hat{m}_{k_l}(\mathbf{x}; \mathbf{H}, 2)\} = \mathbf{e}_{l+1}^{\mathsf{T}} (\mathbf{X}_{2,x}^{\mathsf{T}} \mathbf{W}_{H,x} \mathbf{X}_{2,x})^{-1} \mathbf{X}_{2,x}^{\mathsf{T}} \mathbf{W}_{H,x} \mathbf{m}, \tag{40}$$

recall that $\mathbf{m} = (m(\mathbf{X}_1), \ldots, m(\mathbf{X}_n))^{\mathsf{T}}$. If all third-order partial derivatives of $m(\cdot)$ are continuous in a neighbourhood of $\mathbf{x}$ Taylor expansion gives,

$$\mathbf{m} = \mathbf{X}_{2,x} \begin{pmatrix} m(\mathbf{x}) \\ \nabla_m(\mathbf{x}) \\ \frac{1}{2}\mathrm{vech}\{\mathcal{H}_m(\mathbf{x}) \odot (2\mathbf{1} - \mathbf{I}_{d\times d})\} \end{pmatrix} + \frac{1}{3!} \begin{pmatrix} (d_x^3)m(\mathbf{X}_1 - \mathbf{x}) \\ \cdots \\ (d_x^3)m(\mathbf{X}_n - \mathbf{x}) \end{pmatrix} + \mathbf{R}_m(\mathbf{x}), \tag{41}$$

where $\mathbf{R}_m(\mathbf{x})$ is a vector of Taylor expansion remainder terms. By (40) and (41) the bias is thus given by,

$$\begin{aligned} \mathsf{E}\{\hat{m}_{k_l}(\mathbf{x}; \mathbf{H}, 2) - \frac{\partial m(\mathbf{x})}{\partial x_l}\} &= \mathbf{e}_{l+1}^{\mathsf{T}} (\mathbf{X}_{2,x}^{\mathsf{T}} \mathbf{W}_{H,x} \mathbf{X}_{2,x})^{-1} \mathbf{X}_{2,x}^{\mathsf{T}} \mathbf{W}_{H,x} \\ &\quad \times \left\{ \frac{1}{3!} \begin{pmatrix} (d_x^3)m(\mathbf{X}_1 - \mathbf{x}) \\ \cdots \\ (d_x^3)m(\mathbf{X}_n - \mathbf{x}) \end{pmatrix} + \mathbf{R}_m(\mathbf{x}) \right\}, \end{aligned} \tag{42}$$

Further, let,

$$\begin{aligned} \mathbf{N}_x &= \begin{pmatrix} \nu_{x,11} & \boldsymbol{\nu}_{x,12} & \boldsymbol{\nu}_{x,13} \\ \boldsymbol{\nu}_{x,21} & \boldsymbol{\nu}_{x,22} & \boldsymbol{\nu}_{x,23} \\ \boldsymbol{\nu}_{x,31} & \boldsymbol{\nu}_{x,32} & \boldsymbol{\nu}_{x,33} \end{pmatrix} \\ &\equiv \int_{\mathcal{D}_{x,H}} \begin{pmatrix} 1 \\ \mathbf{u} \\ \mathrm{vech}\{\mathbf{u}\mathbf{u}^{\mathsf{T}}\} \end{pmatrix} \begin{pmatrix} 1 & \mathbf{u}^{\mathsf{T}} & \mathrm{vech}\{\mathbf{u}\mathbf{u}^{\mathsf{T}}\}^{\mathsf{T}} \end{pmatrix} K(\mathbf{u}) d\mathbf{u}, \end{aligned} \tag{43}$$

where $\mathcal{D}_{x,H} = \{\mathbf{z} : (\mathbf{x} + \mathbf{H}\mathbf{z}) \in \mathrm{supp}\{f(\cdot)\}\} \cap \mathrm{supp}\{K(\cdot - \mathbf{x})\}$ is the support of the kernel-function centered at $\mathbf{x}$ that is contained in the support of $f(\cdot)$. Also let the elements of the inverse matrix be denoted by,

$$\mathbf{N}_x^{-1} = \begin{pmatrix} \nu_x^{11} & \boldsymbol{\nu}_x^{12} & \boldsymbol{\nu}_x^{13} \\ \boldsymbol{\nu}_x^{21} & \boldsymbol{\nu}_x^{22} & \boldsymbol{\nu}_x^{23} \\ \boldsymbol{\nu}_x^{31} & \boldsymbol{\nu}_x^{32} & \boldsymbol{\nu}_x^{33} \end{pmatrix}. \tag{44}$$

The notation of the bandwidth matrix is straightforwardly expanded,

$$\mathbf{H} = \begin{pmatrix} h_{x,11} & \dots & h_{x,1d} \\ \vdots & \ddots & \vdots \\ h_{x,1d} & \dots & h_{x,dd} \end{pmatrix}. \tag{45}$$

The following lemma introduces some convenient notation for the entries in $\mathbf{X}_{2,x}^{\mathsf{T}}\mathbf{W}_{H,x}\mathbf{X}_{2,x}$ and provides asymptotic expressions which are collected from Ruppert and Wand (1994). Here $\mathbf{C}$ is the $d(d+1)/2 \times d(d+1)/2$ matrix defined by the relation $\mathrm{vech}(\mathbf{H}\mathbf{u}\mathbf{u}^{\mathsf{T}}\mathbf{H}) = \mathbf{C}\mathrm{vech}(\mathbf{u}\mathbf{u}^{\mathsf{T}})$. The elements of $\mathbf{C}$ contain products of two elements of $\mathbf{H}$ so that they are $O(\lambda_{\max}^2(\mathbf{H}))$.

Lemma 1 *Let,*

$$\begin{aligned}
t_1(\mathbf{x}, n) &\equiv \frac{1}{n}\sum_{i=1}^{n} K_H(\mathbf{X}_i - \mathbf{x}), \\
\mathbf{t}_2(\mathbf{x}, n) &\equiv \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_i - \mathbf{x})\, K_H(\mathbf{X}_i - \mathbf{x}), \\
\mathbf{t}_3(\mathbf{x}, n) &\equiv \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})^{\mathsf{T}}\, K_H(\mathbf{X}_i - \mathbf{x}), \\
\mathbf{t}_4(\mathbf{x}, n) &\equiv \frac{1}{n}\sum_{i=1}^{n}\mathrm{vech}\{(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})^{\mathsf{T}}\}\, K_H(\mathbf{X}_i - \mathbf{x}), \\
\mathbf{t}_5(\mathbf{x}, n) &\equiv \frac{1}{n}\sum_{i=1}^{n}\mathrm{vech}\{(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})^{\mathsf{T}}\}(\mathbf{X}_i - \mathbf{x})^{\mathsf{T}}\, K_H(\mathbf{X}_i - \mathbf{x}), \\
\mathbf{t}_6(\mathbf{x}, n) &\equiv \frac{1}{n}\sum_{i=1}^{n}\mathrm{vech}\{(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})^{\mathsf{T}}\}\mathrm{vech}\{(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})^{\mathsf{T}}\}^{\mathsf{T}}\, K_H(\mathbf{X}_i - \mathbf{x}).
\end{aligned} \tag{46}$$

Assume that (A3) *and* (A4) *hold and that* $f(\cdot)$ *is continuous at* $\mathbf{x}$ *then,*

$$\begin{aligned}
t_1(\mathbf{x}, n) &= \nu_{x,11} f(\mathbf{x}) + o_p(1) \\
\mathbf{t}_2(\mathbf{x}, n) &= \mathbf{H}\boldsymbol{\nu}_{x,21} f(\mathbf{x}) + o_p(\mathbf{H}\mathbf{1}) \\
\mathbf{t}_3(\mathbf{x}, n) &= \mathbf{H}\boldsymbol{\nu}_{x,22}\mathbf{H} f(\mathbf{x}) + o_p(\mathbf{H}\mathbf{1}\mathbf{H}^{\mathsf{T}}) \\
\mathbf{t}_4(\mathbf{x}, n) &= \mathbf{C}\boldsymbol{\nu}_{x,31} f(\mathbf{x}) + o_p(\mathbf{C}\mathbf{1}) \\
\mathbf{t}_5(\mathbf{x}, n) &= \mathbf{C}\boldsymbol{\nu}_{x,32}\mathbf{H} f(\mathbf{x}) + o_p(\mathbf{C}\mathbf{1}\mathbf{H}), \\
\mathbf{t}_6(\mathbf{x}, n) &= \mathbf{C}\boldsymbol{\nu}_{x,33}\mathbf{C}^{\mathsf{T}} f(\mathbf{x}) + o_p(\mathbf{C}\mathbf{1}\mathbf{C}^{\mathsf{T}}).
\end{aligned} \tag{47}$$

Proof of Lemma 1: We give the proof for $\mathbf{t}_6(\mathbf{x}, n)$. The other expressions are proven analogously. Let $\mathbf{t}_6^C(\mathbf{x}, n) = \mathbf{C}^{-1}\, \mathbf{t}_6(\mathbf{x}, n)\, (\mathbf{C}^\mathsf{T})^{-1}$ and it follows that the equation for $\mathbf{t}_6(\mathbf{x}, n)$ is equivalent to $\mathbf{t}_6^C(\mathbf{x}, n) = \boldsymbol{\nu}_{x,33} f(\mathbf{x}) + o_p(\mathbf{1})$. In order to prove this it is sufficient to show that i) $\mathsf{E}\{\mathbf{t}_6^C(\mathbf{x}, n)\} \to \boldsymbol{\nu}_{x,33} f(\mathbf{x})$ as $n \to \infty$ and that ii) $\mathsf{Var}\{\mathbf{t}_6^C(\mathbf{x}, n)\} \to 0$ as $n \to \infty$.

i)

We have,

$$\begin{aligned}
&\mathsf{E}\{\mathbf{t}_6^C(\mathbf{x}, n)\} \\
&= \mathbf{C}^{-1} \int \operatorname{vech}\{(\mathbf{v}-\mathbf{x})(\mathbf{v}-\mathbf{x})^\mathsf{T}\}\operatorname{vech}\{(\mathbf{v}-\mathbf{x})(\mathbf{v}-\mathbf{x})^\mathsf{T}\}^\mathsf{T} \\
&\times |\mathbf{H}|^{-1} K\{\mathbf{H}^{-1}(\mathbf{v}-\mathbf{x})\} f(\mathbf{v})\, d\mathbf{v}\, (\mathbf{C}^\mathsf{T})^{-1}.
\end{aligned} \tag{48}$$

By using the transformation $\mathbf{v} = \mathbf{x} + \mathbf{H}\mathbf{u}$ and Taylor's theorem,

$$\begin{aligned}
&\mathsf{E}\{\mathbf{t}_6^C(\mathbf{x}, n)\} \\
&= \mathbf{C}^{-1} \int_{\mathcal{D}_{x,H}} \operatorname{vech}\{(\mathbf{H}\mathbf{u})(\mathbf{H}\mathbf{u})^\mathsf{T}\}\operatorname{vech}\{(\mathbf{H}\mathbf{u})(\mathbf{H}\mathbf{u})^\mathsf{T}\}^\mathsf{T} K(\mathbf{u}) f(\mathbf{x}+\mathbf{H}\mathbf{u})\, d\mathbf{u}\, (\mathbf{C}^\mathsf{T})^{-1} \\
&= \int_{\mathcal{D}_{x,H}} \operatorname{vech}\{\mathbf{u}\mathbf{u}^\mathsf{T}\}\operatorname{vech}\{\mathbf{u}\mathbf{u}^\mathsf{T}\}^\mathsf{T} K(\mathbf{u})\, d\mathbf{u}\, \{f(\mathbf{x}) + o(1)\} \\
&= \boldsymbol{\nu}_{x,33}\{f(\mathbf{x}) + o(1)\}.
\end{aligned} \tag{49}$$

This shows that $\mathsf{E}\{\mathbf{t}_6^C(\mathbf{x}, n)\} \to \boldsymbol{\nu}_{x,33} f(\mathbf{x})$ as $n \to \infty$.

ii)

Next, we proceed to show that $\mathsf{Var}\{\mathbf{t}_6^C(\mathbf{x}, n)\} \to 0$ as $n \to \infty$. To save space the square of a matrix $\mathbf{A}$, $\mathbf{A}\mathbf{A}^\mathsf{T}$, will be denoted by $\mathbf{A}^2$ here.

$$\begin{aligned}
&\mathsf{Var}\{\mathbf{t}_6^C(\mathbf{x}, n)\} \leq \mathsf{E}\{[\mathbf{t}_6^C(\mathbf{x}, n)]^2\} \\
&= n^{-1} \int [\mathbf{C}^{-1} \operatorname{vech}\{(\mathbf{v}-\mathbf{x})(\mathbf{v}-\mathbf{x})^\mathsf{T}\}\operatorname{vech}\{(\mathbf{v}-\mathbf{x})(\mathbf{v}-\mathbf{x})^\mathsf{T}\}^\mathsf{T} (\mathbf{C}^\mathsf{T})^{-1} \\
&\times |\mathbf{H}|^{-1} K\{\mathbf{H}^{-1}(\mathbf{v}-\mathbf{x})\}]^2 f(\mathbf{v})\, d\mathbf{v} \\
&= n^{-1} \int_{\mathcal{D}_{x,H}} [\mathbf{C}^{-1} \operatorname{vech}\{\mathbf{H}\mathbf{u}\mathbf{u}^\mathsf{T}\mathbf{H}\}\operatorname{vech}\{\mathbf{H}\mathbf{u}\mathbf{u}^\mathsf{T}\mathbf{H}\}^\mathsf{T} (\mathbf{C}^\mathsf{T})^{-1} \\
&\times |\mathbf{H}|^{-1} K(\mathbf{u})]^2 |\mathbf{H}| f(\mathbf{x}+\mathbf{H}\mathbf{u})\, d\mathbf{u} \\
&= n^{-1} \int_{\mathcal{D}_{x,H}} [\operatorname{vech}\{\mathbf{u}\mathbf{u}^\mathsf{T}\}\operatorname{vech}\{\mathbf{u}\mathbf{u}^\mathsf{T}\}^\mathsf{T}]^2\, |\mathbf{H}|^{-1} K^2(\mathbf{u}) f(\mathbf{x}+\mathbf{H}\mathbf{u})\, d\mathbf{u} \\
&= n^{-1}|\mathbf{H}|^{-1} \int_{\mathcal{D}_{x,H}} [\operatorname{vech}\{\mathbf{u}\mathbf{u}^\mathsf{T}\}\operatorname{vech}\{\mathbf{u}\mathbf{u}^\mathsf{T}\}^\mathsf{T}]^2 K^2(\mathbf{u})\, d\mathbf{u}\, \{f(\mathbf{x}) + o(1)\}.
\end{aligned} \tag{50}$$

Since the moments of $K^2(\cdot)$ are bounded it follows from (A3) that $\mathsf{Var}\{\mathbf{t}_6^C(\mathbf{x}, n)\} \to 0$ as $n \to \infty$ since $f(\cdot)$ is bounded and $n^{-1}|\mathbf{H}|^{-1} \to 0$ as $n \to \infty$. This completes the proof. □

Now, consider the asymptotic bias for local quadratic estimation of a first order partial derivate of $m(\cdot)$ at the boundary. Recall that $\mathbf{k}_l$ is a $1 \times d$ vector of zeros except for position l which equals one and let $\tilde{\mathbf{e}}_l$ be the $1 + d + d(d+1)/2 \times 1$ vector of zeros except at locations 2 to $d + 1$ which contains the elements of row l in $\mathbf{H}^{-1}$.

Theorem 2 *Suppose that all third-order partial derivatives of $m(\cdot)$ are continuous in a neighbourhood of $\mathbf{x}_\partial$ on the boundary of* $\text{supp}\{f(\cdot)\}$, $\mathbf{x} = \mathbf{x}_\partial + \mathbf{Hc}$, *where* $\mathbf{c}$ *is a fixed element of* $\text{supp}\{K(\cdot)\}$ *and that $f(\cdot)$ is continuous at* $\mathbf{x}$. *Also, suppose that condition* (A3) *and* (A4) *hold. Then*

$$\begin{aligned}\mathsf{E}\{\hat{m}_{k_l}(\mathbf{x};\mathbf{H},2) - m^{(k)}(\mathbf{x})|\mathbf{X}_1,\ldots,\mathbf{X}_n\} \\ &= \frac{\tilde{\mathbf{e}}_l^{\mathsf{T}} N_x^{-1}}{3!}\int_{\mathcal{D}_{x,H}} \begin{pmatrix} 1 \\ \mathbf{u} \\ \text{vech}\{\mathbf{u}\mathbf{u}^{\mathsf{T}}\}\end{pmatrix} K(\mathbf{u})(d_x^3)m(\mathbf{Hu})d\mathbf{u} \\ &\quad + o_p\{\text{tr}\{\mathbf{H}\}^2\}.\end{aligned} \tag{51}$$

□

Proof of Theorem 2: The result is given by calculation of the asymptotic properties of the bias given in (42). Note that multiplication of $(\mathbf{X}_{2,x}^{\mathsf{T}}\mathbf{W}_{H,x}\mathbf{X}_{2,x})^{-1}\mathbf{X}_{2,x}^{\mathsf{T}}\mathbf{W}_{H,x}$ by $\mathbf{R}_m(\mathbf{x})$ is of negligible order compared to the terms arising from multiplication by the vector containing the third order derivatives. By Lemma 1,

$$n^{-1}\mathbf{X}_{2,x}^{\mathsf{T}}\mathbf{W}_{H,x}\mathbf{X}_{2,x} = \begin{pmatrix} t_1(\mathbf{x},n) & \mathbf{t}_2^{\mathsf{T}}(\mathbf{x},n) & \mathbf{t}_4^{\mathsf{T}}(\mathbf{x},n) \\ \mathbf{t}_2(\mathbf{x},n) & \mathbf{t}_3(\mathbf{x},n) & \mathbf{t}_5^{\mathsf{T}}(\mathbf{x},n) \\ \mathbf{t}_4(\mathbf{x},n) & \mathbf{t}_5(\mathbf{x},n) & \mathbf{t}_6(\mathbf{x},n)\end{pmatrix}, \tag{52}$$

so that,

$$\begin{aligned}n^{-1}\mathbf{X}_{2,x}^{\mathsf{T}}\mathbf{W}_{H,x}\mathbf{X}_{2,x} &= \{\text{diag}(1,\mathbf{H},\mathbf{C})\}\,\mathbf{N}_x\left\{\text{diag}(1,\mathbf{H},\mathbf{C}^{\mathsf{T}})\right\}f(\mathbf{x}) \\ &\quad + o_p\left[\{\text{diag}(1,\mathbf{H},\mathbf{C})\}\,\mathbf{1}\left\{\text{diag}(1,\mathbf{H},\mathbf{C}^{\mathsf{T}})\right\}\right],\end{aligned} \tag{53}$$

where $\text{diag}(1,\mathbf{H},\mathbf{C})$ is a $(1+d+d(d+1)/2)\times(1+d+d(d+1)/2)$ block-diagonal matrix and $\mathbf{1}$ is a generic matrix of ones. The inverse of (53) is thus given by,

$$\begin{aligned}(n^{-1}\mathbf{X}_{2,x}^{\mathsf{T}}\mathbf{W}_{H,x}\mathbf{X}_{2,x})^{-1} &= f^{-1}(\mathbf{x})\left\{\text{diag}(1,\mathbf{H}^{-1},(\mathbf{C}^{\mathsf{T}})^{-1})\right\}\mathbf{N}_x^{-1}\left\{\text{diag}(1,\mathbf{H}^{-1},\mathbf{C}^{-1})\right\} \\ &\quad + o_p\left[\left\{\text{diag}(1,\mathbf{H}^{-1},(\mathbf{C}^{\mathsf{T}})^{-1})\right\}\mathbf{1}\left\{\text{diag}(1,\mathbf{H}^{-1},\mathbf{C}^{-1})\right\}\right].\end{aligned} \tag{54}$$

Next,

$$
\begin{aligned}
& n^{-1}\mathbf{X}_{2,x}^{\mathsf{T}}\mathbf{W}_{H,x}\begin{pmatrix}(d_x^3)m(\mathbf{X}_1-\mathbf{x})\\ \dots\\ (d_x^3)m(\mathbf{X}_n-\mathbf{x})\end{pmatrix}\\
&\quad = f(\mathbf{x})\int_{\mathcal{D}_{x,H}}\begin{pmatrix}1\\ \mathbf{Hu}\\ \mathbf{C}\text{vech}\{\mathbf{uu}^{\mathsf{T}}\}\end{pmatrix}K(\mathbf{u})(d_x^3)m(\mathbf{Hu})d\mathbf{u} \qquad (55)\\
&\quad + o_p\left\{\text{tr}\{\mathbf{H}\}^3\begin{pmatrix}1\\ \mathbf{H1}\\ \mathbf{C1}\end{pmatrix}\right\}.
\end{aligned}
$$

Since this equation is shown in a similar manner to the proof of $\mathbf{t}_6(\mathbf{x}, n)$ given in Lemma 1, we only discuss the lower $d(d+1)/2 \times 1$ block of (55). This block is equal to,

$$
n^{-1}\sum_{i=1}^{n}\text{vech}\{(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})^{\mathsf{T}}\}\,K_H(\mathbf{X}_i-\mathbf{x})(d_x^3)m(\mathbf{X}_i-\mathbf{x}). \qquad (56)
$$

Consider premultiplication of (56) by $\mathbf{C}^{-1}$, then it is sufficient to show that the mean of this sum is $f(\mathbf{x})\int_{\mathcal{D}_{x,H}}\text{vech}\{\mathbf{uu}^{\mathsf{T}}\}K(\mathbf{u})(d_x^3)m(\mathbf{Hu})d\mathbf{u}$ and that the variance converges to zero at the rate $o(\text{tr}\{\mathbf{H}\}^3\mathbf{1})^2$. The mean is straightforwardly shown using analogous calculations as in Lemma 1. The variance is bounded by,

$$
\begin{aligned}
& n^{-1}\int[\mathbf{C}^{-1}\text{vech}\{(\mathbf{v}-\mathbf{x})(\mathbf{v}-\mathbf{x})^{\mathsf{T}}\}\,K_H(\mathbf{v}-\mathbf{x})(d_x^3)m(\mathbf{v}-\mathbf{x})]^2 f(\mathbf{v})d\mathbf{v}\\
& = n^{-1}\int_{\mathcal{D}_{x,H}}[\mathbf{C}^{-1}\text{vech}\{\mathbf{Huu}^{\mathsf{T}}\mathbf{H}\}|\mathbf{H}|^{-1}K(\mathbf{u})(d_x^3)m(\mathbf{Hu})]^2|\mathbf{H}|f(\mathbf{x}+\mathbf{Hu})d\mathbf{u}\\
& = n^{-1}|\mathbf{H}|^{-1}\int_{\mathcal{D}_{x,H}}\text{vech}\{\mathbf{uu}^{\mathsf{T}}\}\text{vech}\{\mathbf{uu}^{\mathsf{T}}\}^{\mathsf{T}}[(d_x^3)m(\mathbf{Hu})]^2K^2(\mathbf{u})d\mathbf{u}\{f(\mathbf{x})+o(1)\}. \qquad (57)
\end{aligned}
$$

In order to find the convergence rate we need to examine the terms of $(d_x^3)m(\mathbf{Hu})$. Since all elements of $\mathbf{H}$ have the same convergence rate, $\mathbf{Hu}$ are $O(tr\{\mathbf{H}\}\mathbf{1})$. Each term of $(d_x^3)m(\mathbf{Hu})$ are $O(tr\{\mathbf{H}\}^3)$ since they contain a product of three elements in $\mathbf{Hu}$. Again, $n^{-1}|\mathbf{H}|^{-1}\to 0$ as $n\to\infty$ so the convergence rate of (57) is $o(\text{tr}\{\mathbf{H}\}^3\mathbf{1})^2$.

By combining (42), (54) and (55) the result is given. □

We now turn to the asymptotic behaviour of the bias in the interior of the support of $f(\cdot)$. For such locations the expression for the asymptotic bias is simpler because some of the blocks in $\mathbf{N}_x$ is zero. Let l and $\mathbf{k}_l$ be defined as above.

Theorem 3 *Suppose that* $\mathbf{x}$ *is in the interior of* $\mathrm{supp}\{f(\cdot)\}$*, that all third-order partial derivatives of* $m(\cdot)$ *are continuous at* $\mathbf{x}$ *and that* $f(\cdot)$ *is continuous at* $\mathbf{x}$*. Also, suppose that condition* (A1) *and* (A3) *hold. Then*

$$\begin{aligned}
&\mathsf{E}\{\hat{m}_{k_l}(\mathbf{x};\mathbf{H},2) - m^{(k)}(\mathbf{x})|\mathbf{X}_1,\ldots,\mathbf{X}_n\} \\
&\qquad = \frac{\mathbf{e}_l^{\mathsf{T}}\mathbf{H}^{-1}}{3!\ \mu_2(K)} \int \mathbf{u}K(\mathbf{u})(d_x^3)m(\mathbf{H}\mathbf{u})d\mathbf{u} \\
&\qquad + o_p\{\mathrm{tr}\{\mathbf{H}\}^2\}.
\end{aligned} \tag{58}$$

□

Proof of Theorem 3: The result given in Theorem 2 also holds for this case, however it is possible to make some simplifications of the expression. Since $\mathbf{x}$ is in the interior of $\mathrm{supp}\{f(\cdot)\}$, $\mathcal{D}_{x,H} = \mathrm{supp}\{K(\cdot)\}$ so that by assumption (A1) $\boldsymbol{\nu}_{x,12}$, $\boldsymbol{\nu}_{x,21}$, $\boldsymbol{\nu}_{x,23}$ and $\boldsymbol{\nu}_{x,32}$ are all zero. Specifically, since $\boldsymbol{\nu}_{x,21}$ and $\boldsymbol{\nu}_{x,23}$ are zero, $\boldsymbol{\nu}_x^{21}$ and $\boldsymbol{\nu}_x^{23}$ are also zero so that

$$\tilde{\mathbf{e}}_l^{\mathsf{T}} N_x^{-1} \begin{pmatrix} 1 \\ \mathbf{u} \\ \mathrm{vech}\{\mathbf{u}\mathbf{u}^{\mathsf{T}}\} \end{pmatrix} = \mathbf{e}_l^{\mathsf{T}}\mathbf{H}^{-1}\boldsymbol{\nu}_x^{22}\,\mathbf{u}. \tag{59}$$

Further $\boldsymbol{\nu}_x^{22} = \mu_2^{-1}(K)\mathbf{I}_d$. This completes the proof. □

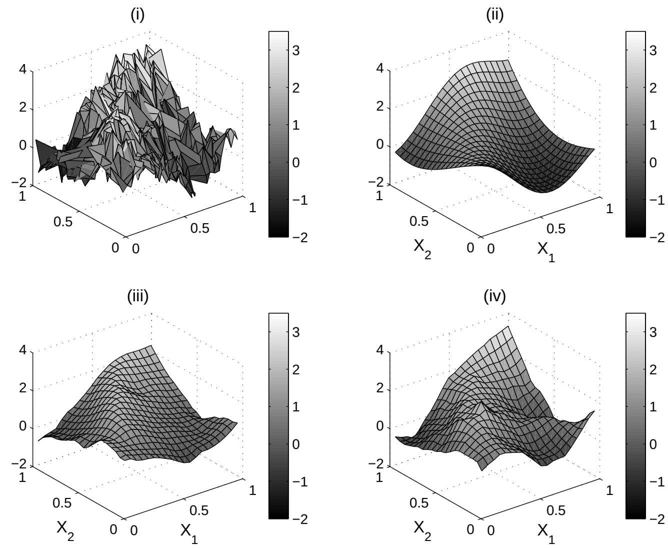

Figure 1: (i) Example of a simulated normal random field following model (1) with mean-function $m(\mathbf{Ax})$ given by (29) and (30), $v(\cdot) = 0.8$ and exponential correlation-function with parameter a=5; (ii) the true mean-function; (iii) $\hat{m}_{0,0}(\cdot; \mathbf{H}, 1)$ using $\text{EBBS}^{\text{full,GM}}_{\text{dep}}$ (iv) $\hat{m}_{0,0}(\cdot; \mathbf{H}, 1)$ using $\text{EBBS}^{\text{full}}$.

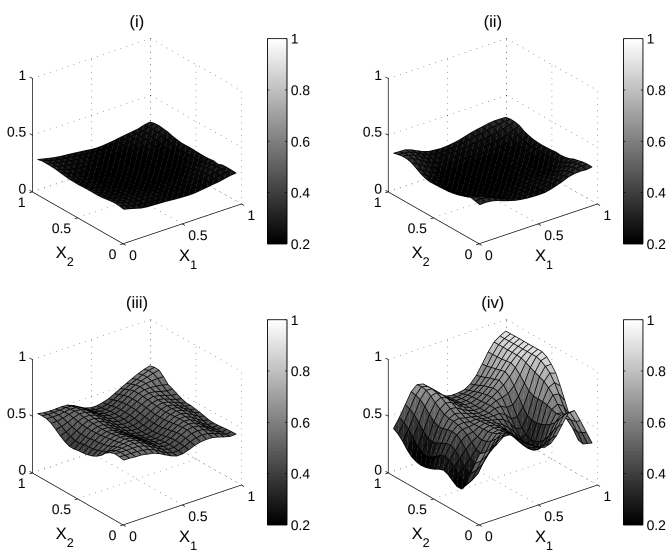

Figure 2: Simulation means of $h_1(\mathbf{x})$ for estimation of $m(\cdot)$ in the model described in Section 3.1 with dependence parameter $a = 5$; (i) $\text{EBBS}^{\text{full}}$; (ii) $\text{EBBS}^{\text{full}}_{\text{dep}}$; (iii) $\text{EBBS}^{\text{full,GM}}_{\text{dep}}$; (iv) $\mathbf{H}_{\text{opt}}$.

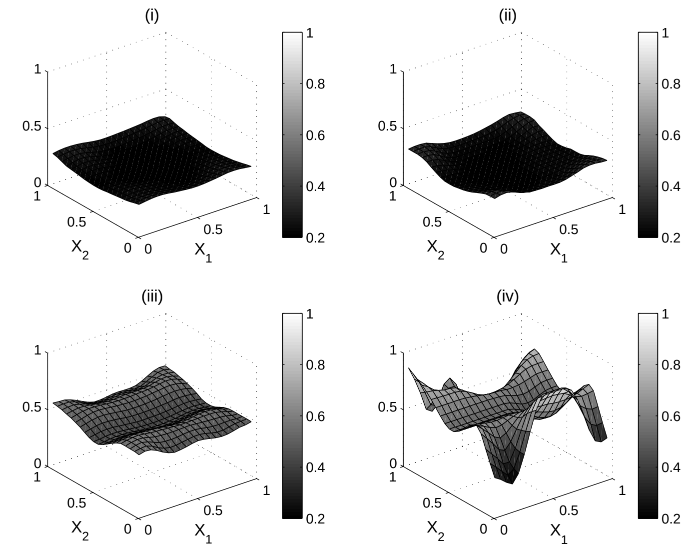

Figure 3: Simulation means of $h_2(\mathbf{x})$ for estimation of $m(\cdot)$ in the model described in Section 3.1 with dependence parameter $a = 5$; (i) $\text{EBBS}^{\text{full}}$; (ii) $\text{EBBS}^{\text{full}}_{\text{dep}}$; (iii) $\text{EBBS}^{\text{full,GM}}_{\text{dep}}$; (iv) $\mathbf{H}_{\text{opt}}$.

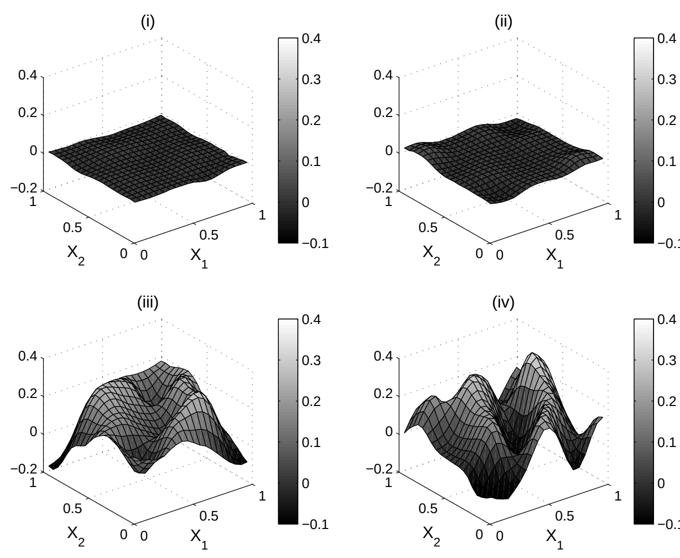

Figure 4: Simulation means of $h_{12}(\mathbf{x})$ for estimation of $m(\cdot)$ in the model described in Section 3.1 with dependence parameter $a = 5$; (i) $\text{EBBS}^{\text{full}}$; (ii) $\text{EBBS}^{\text{full}}_{\text{dep}}$; (iii) $\text{EBBS}^{\text{full,GM}}_{\text{dep}}$; (iv) $\mathbf{H}_{\text{opt}}$.

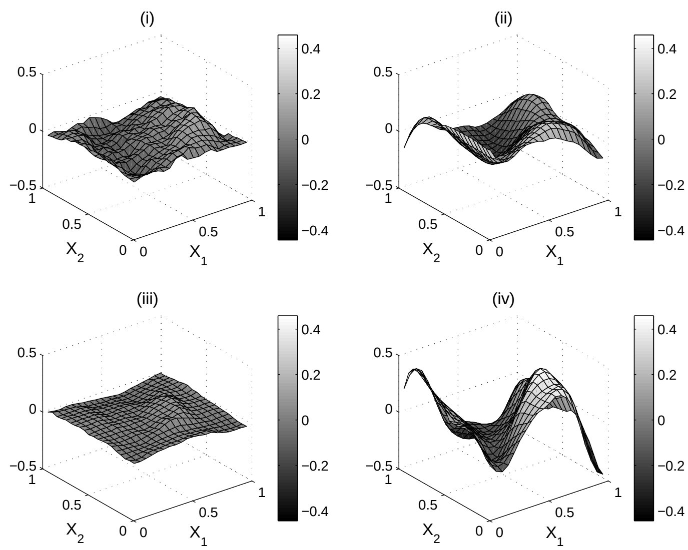

Figure 5: Performance of the empirical bias estimate for the model with dependence parameter $a = 5$; (i) mean of the estimated bias by $\mathrm{EBBS}^{\mathrm{full}}_{\mathrm{dep}}$; (ii) simulated mean of the bias by $\mathrm{EBBS}^{\mathrm{full}}_{\mathrm{dep}}$; (iii) mean of the estimated bias by $\mathrm{EBBS}^{\mathrm{full,GM}}_{\mathrm{dep}}$; (iv) simulated mean of the bias by $\mathrm{EBBS}^{\mathrm{full,GM}}_{\mathrm{dep}}$.

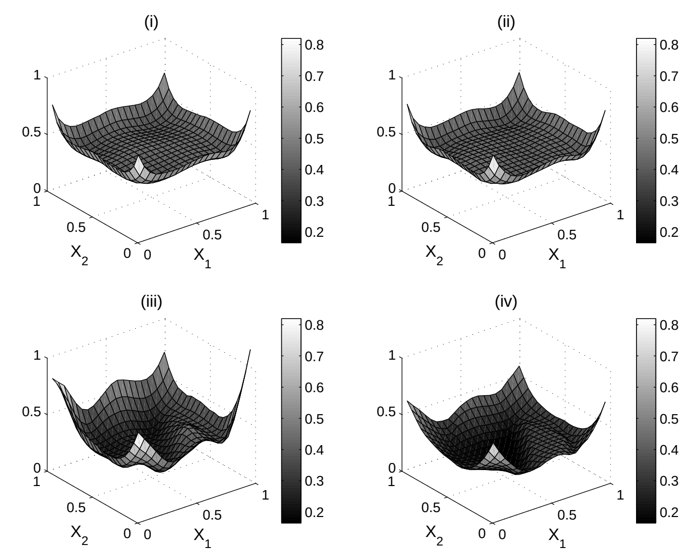

Figure 6: Means of the MSE fields in the simulations for $a = 5$ using; (i) $\mathrm{EBBS}^{\mathrm{full}}$; (ii) $\mathrm{EBBS}^{\mathrm{full}}_{\mathrm{dep}}$; (iii) $\mathrm{EBBS}^{\mathrm{full,GM}}_{\mathrm{dep}}$ (iv) MSE for the optimal bandwidths.

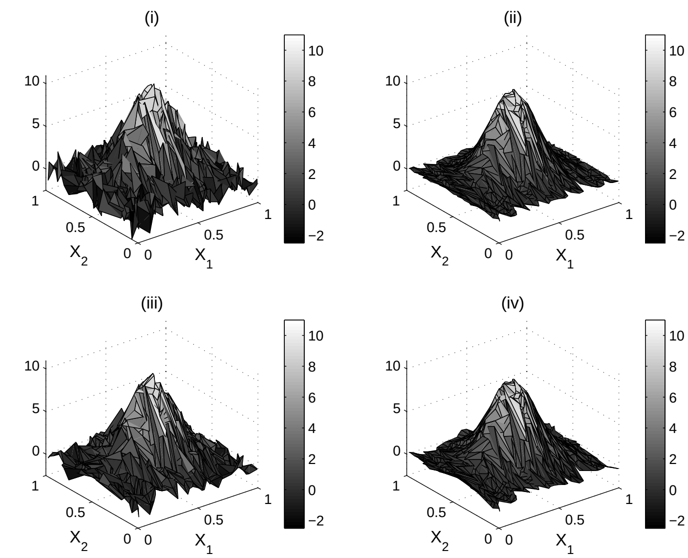

Figure 7: (i) Realization of a normal random field with the mean-function specified in (34), $v(\cdot) = 1$ and iid errors; (ii) $m(\cdot)$; (iii) $\hat{m}(\cdot)$ using span (0.15, 0.15); (iv) $\hat{m}(\cdot)$ using span (0.40, 0.40).

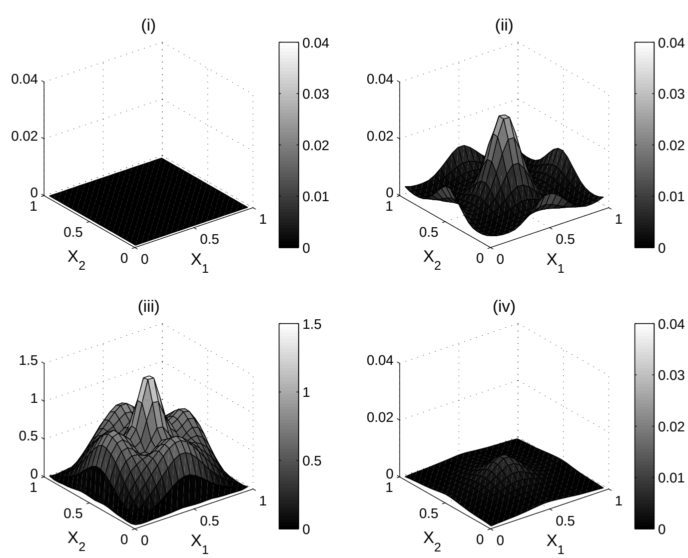

Figure 8: Mean of the true bias of $\hat{v}(\cdot)$ when initial estimate of $\hat{m}(\cdot)$ uses; (i) span (0.15, 0.15); (ii) span (0.40, 0.40) (iii) span (0.75, 0.75), (note the different scale) (iv) EBBS_{dep}.

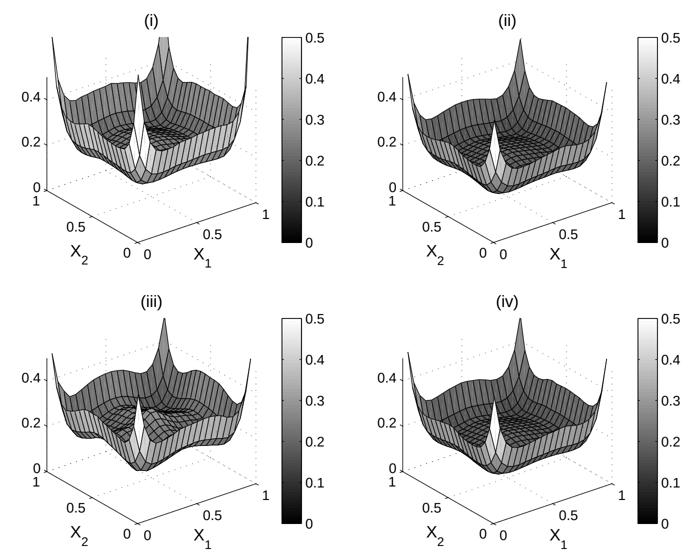

Figure 9: Mean of the true standard deviation of $\hat{v}(\cdot)$ when initial estimate of $\hat{m}(\cdot)$ uses; (i) span (0.15, 0.15); (ii) span (0.40, 0.40) (iii) span (0.75,0.75) (iv) EBBS_{dep}.

MASE, $\mathbf{\Sigma}$

a	EBBS	$\mathrm{EBBS}_{\mathrm{dep}}$	$\mathrm{EBBS}_{\mathrm{dep}}^{\mathrm{GM}}$	$\mathbf{H}_{\mathrm{opt}}$
5	0.4573	0.4329	0.3813	0.3452
20	0.1554	0.1369	0.1268	0.0905
40	0.0702	0.0644	0.0752	0.0425
200	0.0267	0.0265	0.0405	0.0221

Table 1: Means of the MASE's corresponding to the different levels of dependence for EBBS, $\mathrm{EBBS}_{\mathrm{dep}}$ and $\mathrm{EBBS}_{\mathrm{dep}}^{\mathrm{GM}}$ using diagonal bandwidth matrices as well as the optimal MASE.

MASE, $\mathbf{\Sigma}$

a	$\mathrm{EBBS}^{\mathrm{full}}$	$\mathrm{EBBS}_{\mathrm{dep}}^{\mathrm{full}}$	$\mathrm{EBBS}_{\mathrm{dep}}^{\mathrm{full,GM}}$	$\mathbf{H}_{\mathrm{opt}}^{\mathrm{full}}$
5	0.4434	0.4402	0.3784	0.3304
20	0.1481	0.1370	0.1516	0.0871
40	0.0663	0.0623	0.0863	0.0415
200	0.0267	0.0265	0.0500	0.0238

Table 2: Means of the MASE's corresponding to the different levels of dependence for $\mathrm{EBBS}^{\mathrm{full}}$, $\mathrm{EBBS}_{\mathrm{dep}}^{\mathrm{full}}$ and $\mathrm{EBBS}_{\mathrm{dep}}^{\mathrm{full,GM}}$ using full bandwidth matrices as well as the optimal MASE.

a	$\hat{a}$	$\hat{b}$
5	10.8205	0.5137
20	20.1572	0.8140
40	34.3491	0.8625
200	146.8464	0.8918

Table 3: Means of the estimated parameters of the exponential covariance-function.

MASE, $\widehat{\boldsymbol{\Sigma}}$

a	EBBS	EBBS_{dep}	$\text{EBBS}_{\text{dep}}^{\text{GM}}$	$\mathbf{H}_{\text{opt}}$
5	0.4663	0.4353	0.3846	0.3452
20	0.1546	0.1367	0.1275	0.0905
40	0.0721	0.0657	0.0756	0.0425
200	0.0271	0.0269	0.0431	0.0221

Table 4: Means of the MASE's corresponding to the different levels of dependence for EBBS, EBBS_{dep} and $\text{EBBS}_{\text{dep}}^{\text{GM}}$ using diagonal bandwidth matrices as well as the optimal MASE.

MASE, $\widehat{\boldsymbol{\Sigma}}$

a	$\text{EBBS}^{\text{full}}$	$\text{EBBS}_{\text{dep}}^{\text{full}}$	$\text{EBBS}_{\text{dep}}^{\text{full,GM}}$	$\mathbf{H}_{\text{opt}}^{\text{full}}$
5	0.4548	0.4374	0.4197	0.3304
20	0.1468	0.1344	0.1501	0.0871
40	0.0682	0.0641	0.0971	0.0415
200	0.0270	0.0269	0.0617	0.0238

Table 5: Means of the MASE's corresponding to the different levels of dependence for $\text{EBBS}^{\text{full}}$, $\text{EBBS}_{\text{dep}}^{\text{full}}$ and $\text{EBBS}_{\text{dep}}^{\text{full,GM}}$ using full bandwidth matrices as well as the optimal MASE.

Span	(0.15, 0.15)	(0.40, 0.40)	(0.75, 0.75)	EBBS_{dep}
MASE	0.0759	0.0475	0.3666	0.0487

Table 6: Mean average squared errors for the variance-function estimates based on three different choices of constant-span initial bandwidths as well as initial bandwidths based on EBBS_{dep}.

References

G. Christakos. On the problem of permissible covariance and variogram models. *Water Resources Research*, 20(2):251–265, 1984.

N. Cressie. *Statistics for Spatial Data*. John Wiley & Sons, revised edition, 1993.

J. Fan, T. Gasser, I Gijbels, M Brockmann, and J. Engel. Local polynomial regression: Optimal kernels and asymptotic minimax efficiency. *Annals of the Institute of Statistical Mathematics*, 49(1):79–99, 1997.

J. Fan and I. Gijbels. *Local Polynomial Modelling and Its Applications*. Chapman & Hall, 1996.

M. Francisco-Fernández and J. D. Opsomer. Smoothing parameter selection methods for nonparametric regression with spatially correlated errors. *Department of Statistics, Iowa State University. Submitted to Biometrics*, (Preprint #03-07), 2003.

P. Hall, N. I. Fisher, and B. Hoffman. On the nonparametric estimation of covariance functions. *The Annals of Statistics*, 22(4):2115–2134, 1994.

Björn Holmquist. Moments and cumulants of the multivariate normal distribution. *Stochastic Analysis and Applications*, 6(3):273–278, 1988.

T. Lindström. A nonparametric covariance estimator for spatial models. Technical Report 2003:02, Centre for Mathematical Sciences, Mathematical Statistics, 2003.

T. Lindström, U. Holst, and P. Weibring. Analysis of lidar fields using local polynomial regression. Technical Report 2004:12, Centre for Mathematical Sciences, Mathematical Statistics. Submitted to Environmetrics, 2004.

Xiao-Hu Liu. *Kernel smoothing for spatially correlated data*. PhD thesis, Iowa State University, 2001.

J. Opsomer, Y. Wang, and Y. Yang. Nonparametric Regression with Correlated Errors. *Statistical Science*, 16(2):134–153, 2001.

J. D. Opsomer, D. Ruppert, M. P. Wand, U. Holst, and O. Hössjer. Kriging with Nonparametric Variance Function Estimation. *Biometrics*, 55:704–710, 1999.

D. Ruppert. Empirical-Bias Bandwidths for Local Polynomial Nonparametric Regression and Density Estimation. *Journal of the American Statistical Association*, 92 (439):1049–1062, 1997.

D. Ruppert and M. P. Wand. Multivariate locally weighted least squares regression. *The Annals of Statistics*, 22(3):1346–1370, 1994.

D. Ruppert, M. P. Wand, U. Holst, and O. Hössjer. Local Polynomial Variance-Function Estimation. *Technometrics*, 39(3):262–273, 1997.

A. Shapiro and J. D. Botha. Variogram fitting with a general class of conditionally nonnegative definite functions. *Computational Statistics & Data Analysis*, 11:87–96, 1991.

Preprints in Mathematical Sciences 2004:11
ISSN 1403-9338

LUTFMS-5035-2004

Mathematical Statistics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

http://www.maths.lth.se/