A goodness-of-fit test for regression models with spatially correlated errors

Andrea Meilán-Vila · Jean D. Opsomer · Mario Francisco-Fernández · Rosa M. Crujeiras

Abstract The problem of assessing a parametric regression model in the presence of spatial correlation is addressed in this work. For that purpose, a goodness-of-fit test based on a L_2-distance comparing a parametric and a nonparametric regression estimators is proposed. Asymptotic properties of the test statistic, both under the null hypothesis and under local alternatives, are derived. Additionally, a bootstrap procedure is designed to calibrate the test in practice. Finite sample performance of the test is analyzed through a simulation study, and its applicability is illustrated using a real data example.

Keywords Model checking · Spatial correlation · Local linear regression · least squares · Bootstrap

Mathematics Subject Classification (2010) 62G10 · 62H11 · 62G08 · 62G09

Andrea Meilán-Vila
Research group MODES, Department of Mathematics, Faculty of Computer Science, Universidade da Coruña, Campus de Elviña s/n, 15071, A Coruña, Spain
Tel.: +34981167000 (ext 1301)
Fax: +34981167160
E-mail: andrea.meilan@udc.es

Jean D. Opsomer
Westat, 1600 Research Boulevard Rockville, MD, 20850, Maryland, USA

Mario Francisco-Fernández
Research group MODES, Department of Mathematics, Faculty of Computer Science, Universidade da Coruña, Campus de Elviña s/n, 15071, A Coruña, Spain

Rosa M. Crujeiras
Department of Statistics, Mathematical Analysis and Optimization, Faculty of Mathematics, Universidade de Santiago de Compostela, Rúa Lope Gómez de Marzoa s/n, 15782, Santiago de Compostela, Spain

1 Introduction

The problem of testing a parametric regression model, confronting a parametric estimator of the regression function with a smooth alternative estimated by a nonparametric method, has been approached by several authors in the statistical literature (see, for example Azzalini et al., 1989; Eubank and Spiegelman, 1990). For instance, Weihrather (1993) and Eubank et al. (2005) described tests based on an overall distance between parametric and nonparametric regression fits, giving some strategies on bandwidth selection. Härdle and Mammen (1993) proposed a testing procedure to check if a regression function belongs to a class of parametric models by measuring a L_2-distance between parametric and nonparametric estimates. Specifically, the Nadaraya-Watson estimator (Nadaraya, 1964; Watson, 1964) was considered for the nonparametric approach. The same type of study was performed by Alcalá et al. (1999), but using a local polynomial regression estimator (Fan and Gijbels, 1996). Following similar ideas, a local test for a univariate parametric model checking was proposed by Opsomer and Francisco-Fernández (2010), while Li (2005) assessed the lack of fit of a nonlinear regression model, comparing a local linear smoother and parametric fits.

The previous testing procedures, all of them formulated with independent errors, have been also adapted for scenarios where data exhibit correlation in time. For example, Park et al. (2015) considered a model specification test based on a kernel for a nonparametric regression model with an equally-spaced fixed design and correlated errors. Also in the context of time series, goodness-of-fit tests for linear regression models with correlated errors have been studied by González-Manteiga and Vilar-Fernández (1995), also considering an equispaced fixed design. Biedermann and Dette (2000) extended the previous results under fixed alternatives, considering a regression model with explanatory variables x_i, $i = 1, \ldots, n$, being fixed and given by $i/n = \int_0^{x_i} f(t)dt$, where f is a positive density on the interval $[0, 1]$. For further discussion and examples of nonparametric specification tests for regression models, see the comprehensive review by González-Manteiga and Crujeiras (2013).

Although for time dependent errors, the problem of assessing a parametric regression model has been widely studied, this is not the case for spatial (or even spatio-temporal) correlated data. Observations from a spatially varying processes are quite frequent in applied sciences such as ecology, environmental and soil sciences. In order to gain some insight in the process evolution accross space, a regression model where the regression function captures the first-order structure, whereas the error term collects the second-order structure, can be formulated in the previous contexts. Usually, parametric models are considered for the regression function, e.g. polynomial models on latitude and longitude (see Cressie, 1993; Diggle and Ribeiro, 2007), and estimation is accomplished by least squares methods, providing reliable inferences if the model is correctly specified. As an example, a classical dataset which is analyzed under this scope is the Wolfcamp aquifer data presented by Harper and Furr (1986), collecting 85 measurements of levels of piezometric-head. In this example, several parametric trend models are considered after performing different analyses, concluding that a linear trend seems to be a reasonable model (see Figure 1). However, to determine if this linear model (or in general, any parametric fit) is an appropriate representation of a dataset, it would be advisable to carry out a statistical test in order to assess the goodness-of-fit of the selected model. In this context, the statistical literature initially focused on the assessment of independence (Diblasi and Bowman, 2001) and on testing a parametric correlation model (Maglione and Diblasi, 2004), considering the variogram as the function describing the spatial dependence pattern. Also taking the variogram as the target function, Bowman and Crujeiras (2013) proposed some testing methods for simplifying hypothesis (namely, stationarity and isotropy). Although these proposals investigate the dependence structure of the data (a nuisance when the primary goal is the regression or trend function), the ideas which inspired these methods are common to the goodness-of-fit tests for regression models.

A new proposal for testing a parametric regression model (with univariate responses and possibly d-dimensional covariates), in the presence of spatial correlation, is presented in this work. Following similar ideas as those of Härdle and Mammen (1993), the test statistic is based on a comparison between a smooth version of a parametric fit and a nonparametric estimator of the regression function, using a weighted L_2-distance. The null hypothesis that the regression function follows a parametric model is rejected if the distance exceeds a certain threshold. To perform the parametric estimation, an iterative procedure based on generalized least squares is used (see Diggle and Ribeiro, 2007), although other fitting techniques such as maximum likelihood methods could be employed. For the nonparametric alternative, the multivariate local linear regression estimator is used (Liu, 2001; Francisco-Fernandez and Opsomer, 2005; Hallin et al., 2004), generalizing in some way the results of Alcalá et al. (1999) for the univariate case with independent errors.

This paper is organized as follows. Section 2 introduces the regression model, as well as the nonparametric and parametric estimators of the regression function used in our approach. Assumptions and the asymptotic distribution of the proposed test statistic, jointly with a bootstrap procedure to calibrate the test are presented in Section 3. A simulation study for assessing the final performance of the test is provided in Section 4. Finally, Section 5 shows how to apply the testing procedure to the Wolfcamp aquifer dataset introduced above. Supplementary materials with the detailed proofs and further simulation results are also available.

2 Statistical model

Denote by $\{(\mathbf{X}_i, Z_i)\}_{i=1}^n$ a random sample of $(d+1)$-valued random vectors, where Z_i denotes a scalar response which depends on a d-dimensional covariate $\mathbf{X}$, with support $D \subset \mathbb{R}^d$, through the following regression model:

$$Z_i = m(\mathbf{X}_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where m is the regression function and ε denotes a spatially correlated error process, which is assumed to be second order stationary, where

$$\mathbb{E}[\varepsilon_i] = 0, \quad \mathrm{Cov}(\varepsilon_i, \varepsilon_j) = \Sigma(i,j) = \sigma^2 \rho_n(\mathbf{X}_i - \mathbf{X}_j), \quad i,j = 1,\ldots,n,$$

with σ^2 being the point variance and ρ_n a continuous stationary correlation function satisfying $\rho_n(0) = 1$, $\rho_n(\mathbf{x}) = \rho_n(-\mathbf{x})$, and $|\rho_n(\mathbf{x})| \leq 1$, $\forall \mathbf{x}$. The subscript n in ρ_n allows the correlation function to shrink as $n \to \infty$ (this will be made more precise below). Under these assumptions, the semivariogram function γ_n satisfies that $\gamma_n(\mathbf{u}) = \sigma^2(1 - \rho_n(\mathbf{u}))$, $\forall \mathbf{u} \in \mathbb{R}^d$. For simplicity, the subscript n will be sometimes omitted. It should be noted that the previous expression for the covariance of the errors is correct if the nugget effect, denoted by c_0, is equal to zero. If $c_0 \neq 0$, then $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = c_1 \rho_n(\mathbf{X}_i - \mathbf{X}_j)$, if $i \neq j$, where $c_1 = \sigma^2 - c_0$ is the partial sill. In what follows, only the case of $c_0 = 0$ is considered. However, the case of considering a nugget effect has also been analyzed through simulations.

The goal of this work is to propose and study a testing procedure to assess the goodness-of-fit of a parametric regression model, that is:

$$H_0 : m \in \mathcal{M}_{\boldsymbol{\beta}} = \{m_{\boldsymbol{\beta}}, \boldsymbol{\beta} \in \mathcal{B}\}, \qquad \text{vs.} \qquad H_a : m \notin \mathcal{M}_{\boldsymbol{\beta}}, \tag{2}$$

where $\mathcal{B} \subset \mathbb{R}^p$ is a compact set, and p denotes the dimension of the parameter space $\mathcal{B}$. For example, in the bidimensional case ($d = 2$), considering that $\mathcal{M}_{\boldsymbol{\beta}}$ is the family of linear models, then $p = 3$. In addition, $m_{\boldsymbol{\beta}}$ denotes a d-variate parametric function with parameter vector $\boldsymbol{\beta}$. Note that $m_{\boldsymbol{\beta}}$ is not restricted to be polynomial, although that is a common choice in practice.

As pointed out in the Introduction, the goodness-of-fit test is based on a weighted L_2-distance which measures the discrepancy between a smooth version of a parametric estimator and a nonparametric estimator of the regression function. H_0 is rejected if the distance between both fits exceeds a critical value. The estimation methods (parametric and nonparametric) considered in this proposal will be described below. As it will be seen in Section 3, the parametric estimator which is used in the test must satisfy a $\sqrt{n}-$consistency property. As an example, an iterative least squares estimator will be also presented.

A note of caution should be made about regression estimation in this context: for spatially correlated data, when just a single realization of the process $\mathrm{F}\{z_1,\ldots,z_n\}$ is available, additional stationarity assumptions on the process are required in order to enable statistical inference. In addition, it should be also noted that, from a single realization, it may be difficult to disentangle the regression and error components, specially if the dependence is strong.

2.1 Nonparametric regression estimation

For the nonparametric estimation of m in model (1), the multivariate local linear estimator (Fan and Gijbels, 1996) is employed. This nonparametric approach presents some advantages over other kernel-type methods (Hallin et al., 2004). For example, it adapts to a broad class of design densities. Moreover, unlike other kernel-type smoothers, this estimator does not suffer from boundary effects. In the spatial framework, the local linear estimator for m at a location $\mathbf{x}$ can be explicitly written as:

$$\hat{m}^{LL}_{\mathbf{H}}(\mathbf{x}) = \mathbf{e}_1'(X_x'W_xX_x)^{-1}X_x'W_x\mathbf{Z}, \tag{3}$$

where $\mathbf{e}_1$ is a vector of length $(d+1)$ with value 1 in the first entry and all other entries 0, X_x is a $n \times (d+1)$ matrix with i-th row equal to $(1, (\mathbf{X}_i - \mathbf{x})')$, $W_x = \text{diag}\{K_{\mathbf{H}}(\mathbf{X}_1 - \mathbf{x}), \ldots, K_{\mathbf{H}}(\mathbf{X}_n - \mathbf{x})\}$, with $K_{\mathbf{H}}(\mathbf{x}) = |\mathbf{H}|^{-1}K(\mathbf{H}^{-1}\mathbf{x})$, being K a d-dimensional kernel function and $\mathbf{H}$ a $d \times d$ symmetric positive definite matrix, and $\mathbf{Z} = (Z_1, \ldots, Z_n)'$.

For the case of uncorrelated data with a random design, Ruppert and Wand (1994) derived the asymptotic mean squared error (AMSE) formula for the multivariate local linear estimator, while Liu (2001) generalized those results when the errors are correlated. The bandwidth matrix $\mathbf{H}$ controls the shape and the size of the local neighborhood used to estimate $m(\mathbf{x})$ and its selection plays an important role in the estimation process. If $\mathbf{H}$ is "small" an undersmoothed estimator is obtained with high variability and, on the other hand, if $\mathbf{H}$ is "large", the resulting estimator will be very smooth and possible with larger bias. Cross-validation procedures for bandwidth selection are the usual ones in this context, but this type of methods derived under independence should not be used directly when data exhibit dependence given that its expectation is severely affected by the correlation (Liu, 2001). In that case, the dependence of the observations should be taken into account in some way in the bandwidth selection method to estimate "optimal" smoothing parameters (Liu, 2001; Francisco-Fernandez and Opsomer, 2005).

2.2 Parametric regression estimation

As pointed out previously, the goodness-of-fit test proposed in this paper also requires of a parametric estimation of the regression function. As it will be remarked in the next section, the test statistic can be applied taking any parametric estimator, as long as it satisfies a consistency property. Specifically, if $m_{\boldsymbol{\beta}_0}$ denotes the "true" regression function under the null hypothesis, and $m_{\hat{\boldsymbol{\beta}}}$ the corresponding parametric estimator, it is needed that the difference $m_{\hat{\boldsymbol{\beta}}}(\mathbf{x}) - m_{\boldsymbol{\beta}_0}(\mathbf{x}) = O_p(n^{-1/2})$ uniformly in $\mathbf{x}$. A suitable parametric estimator satisfying this property is, for example, the one considered by Crujeiras and Van Keilegon (2010), and this is the parametric method employed for the practical application of the test.

The parametric estimator studied by Crujeiras and Van Keilegon (2010) is obtained using an iterative least squares algorithm. A feasible version of this method includes an approximation of the variance-covariance matrix of the errors. However, for estimating the covariance structure, an initial estimation of the regression is required. This feature leads to the design of iterative estimation procedures in this setting. Following these ideas, this parametric regression estimator is computed as follows:

1. Get an initial estimator of $\boldsymbol{\beta}$ by least squares regression:
$$\tilde{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} (\mathbf{Z} - \mathbf{m}_{\boldsymbol{\beta}})'(\mathbf{Z} - \mathbf{m}_{\boldsymbol{\beta}}), \tag{4}$$
where $\mathbf{m}_{\boldsymbol{\beta}} = (m_{\boldsymbol{\beta}}(\mathbf{X}_1), \ldots, m_{\boldsymbol{\beta}}(\mathbf{X}_n))'$ is the regression function evaluated at the explicative variables.
2. Using the residuals $\tilde{\varepsilon}_i = Z_i - m_{\tilde{\boldsymbol{\beta}}}(\mathbf{X}_i)$, $i = 1, \ldots, n$, and assuming that the variogram belongs to a valid parametric family $\{2\gamma_{\boldsymbol{\theta}},\ \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^q\}$ (usually $q = 3$, with the vector $\boldsymbol{\theta}$ made up of the nugget effect, the partial sill, and the practical range), obtain a parameter estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$. Following a classical approach, $\boldsymbol{\theta}$ is approximated by fitting the parametric model considered for the variogram to a set of empirical semivariogram estimates, computed using the residuals $\tilde{\varepsilon}_i$, applying the weighted least squares method (Cressie, 1985). Under this parametric assumption, the variance-covariance matrix of the errors can be denoted by $\Sigma_{\boldsymbol{\theta}}$, with elements $\Sigma_{\boldsymbol{\theta}}(i,j)$, $i, j = 1 \ldots, n$. Then, replacing $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}$ in these elements, a parametric estimation of $\Sigma_{\boldsymbol{\theta}}$ (denoted by $\Sigma_{\hat{\boldsymbol{\theta}}}$) is obtained.
3. Using $\Sigma_{\hat{\boldsymbol{\theta}}}$, estimate the regression parameter $\boldsymbol{\beta}$ applying the weighted least squares method:
$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} (\mathbf{Z} - \mathbf{m}_{\boldsymbol{\beta}})'\Sigma_{\hat{\boldsymbol{\theta}}}^{-1}(\mathbf{Z} - \mathbf{m}_{\boldsymbol{\beta}}). \tag{5}$$
Finally, the parametric estimator of m considered is given by $m_{\hat{\boldsymbol{\beta}}}$.

3 Test statistic

As pointed out in Section 2, the aim of this paper is to propose a goodness-of-fit test to check if the regression function in model (1) can be assumed to belong to a certain parametric family, $\{m_{\boldsymbol{\beta}}, \boldsymbol{\beta} \in \mathcal{B}\}$. To tackle this problem, a natural approach consists in comparing a parametric estimator of the regression function with a nonparametric one. The question arises if the differences between both fits can be explained by small stochastic fluctuations or if such differences suggest that the parametric assumption is not correct and it is more reasonable to use nonparametric methods to approximate the regression function. Using these ideas, one way to proceed is to measure the distance between both fits and to employ this distance as the test statistic for checking the parametric model.

The approach followed in this work to solve this problem, as in Alcalá et al. (1999), considers a test statistic given by a weighted L_2-distance between the nonparametric and parametric fits to address the testing problem (2):
$$T_n = n|\mathbf{H}|^{1/2} \int_D (\hat{m}^{LL}_{\mathbf{H}}(\mathbf{x}) - \hat{m}^{LL}_{\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x}))^2 w(\mathbf{x}) d\mathbf{x}, \tag{6}$$

where w is a weight function that helps in mitigating possible edge effects. The use of a weight function is quite frequent in this type of tests, both for density and regression (González-Manteiga and Crujeiras, 2013). Moreover, $\hat{m}^{LL}_{\mathbf{H},\hat{\boldsymbol{\beta}}}$ is a smooth version of the parametric estimator $m_{\hat{\boldsymbol{\beta}}}$ which is defined by

$$\hat{m}^{LL}_{\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x}) = \mathbf{e}_1'(X_x'W_xX_x)^{-1}X_x'W_x\mathbf{m}_{\hat{\boldsymbol{\beta}}}, \tag{7}$$

with $\mathbf{m}_{\hat{\boldsymbol{\beta}}} = (m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_1), \ldots, m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_n))'$.

In the particular situation that the parametric family $\mathcal{M}_{\boldsymbol{\beta}}$ in (2) is the class of polynomials of degree less or equal than k, it could be more reasonable to use, as the nonparametric fit, the multivariate local polynomial estimator of degree l, with $l \geq k$, and considering the L_2-distance between this estimator and $m_{\hat{\boldsymbol{\beta}}}$. In that case, it would not be necessary to employ a smooth version of $m_{\hat{\boldsymbol{\beta}}}$, because both are consistent unbiased estimators of the regression function, under the null hypothesis. However, for a general parametric family $\mathcal{M}_{\boldsymbol{\beta}}$, this is not true, and using the simpler local linear estimator, given that $\mathbb{E}[\hat{m}^{LL}_{\mathbf{H}}(\mathbf{x})] = \mathbf{e}_1'(X_x'W_xX_x)^{-1}X_x'W_xm(\mathbf{x})$, it is convenient to smooth the parametric estimator so that the parametric term in (6) has the same expected value as the nonparametric term, under H_0. This fact also justifies the use of the same bandwidth matrix $\mathbf{H}$ in $\hat{m}^{LL}_{\mathbf{H}}$ and in $\hat{m}^{LL}_{\mathbf{H},\hat{\boldsymbol{\beta}}}$ (see Härdle and Mammen, 1993, p. 1928). It is clear that the statistic T_n will be large when the parametric and nonparametric fits, evaluated on the domain D, differ substantially.

For example, considering the Wolfcamp aquifer dataset described in the Introduction, Figure 2 shows the smooth version of the parametric (left) and the nonparametric (right) regression estimators for the level of piezometric-head in the area of study. In this case, a linear model is considered for the parametric fit, while the local linear estimator (3) is employed to perform the nonparametric fit (specific details on the estimation procedures and the fits will be discussed later). Given that both surfaces are very similar, the value of the test statistic T_n will be *small*, and there may be no evidences against the assumption of a linear trend. This feature will indeed be confirmed with the statistical illustration of (6) presented in Section 5.

The types of model deviations that can be captured by this test are of the form $m(\mathbf{x}) = m_{\boldsymbol{\beta}_0}(\mathbf{x}) + c_ng(\mathbf{x})$, where c_n is a sequence, such that $c_n \to 0$ and g is a deterministic function collecting the deviation direction from the null model. In the following section, the asymptotic distribution of the test statistic (6) is derived under the null hypothesis, and also under local alternatives converging to the null hypothesis at a certain rate controlled by c_n. Specifically, it is assumed that the function g is bounded (uniformly in $\mathbf{x}$ and n) and $c_n = n^{-1/2}|\mathbf{H}|^{-1/4}$. In particular, this contains the null hypothesis corresponding to $g(\mathbf{x}) = 0$.

It is clear from expression (6) that T_n depends on the bandwidth matrix $\mathbf{H}$. While the bandwidth selection problem has been well studied in the regression estimation framework, it is still an open issue in goodness-of-fit studies relying on nonparametric methods. In this paper, the smoothing parameter selection problem is not investigated further. Instead, the performance of the test statistic T_n is analyzed for a range of bandwidths in the numerical studies, allowing to check how sensitive the results are to variations in $\mathbf{H}$. Note that although technically it is possible to consider different bandwidth matrices in $\hat{m}_{\mathbf{H}}^{LL}$ and $\hat{m}_{\mathbf{H},\hat{\boldsymbol{\beta}}}^{LL}$, the use of just one bandwidth matrix simplifies the application of the test in practice.

Note that the test statistic (6) generalizes to the framework of spatial correlated data (with a d-dimensional covariate) the statistic proposed for independent data by Härdle and Mammen (1993), using the Nadaraya-Watson estimator, and that of Alcalá et al. (1999) using the local polynomial estimator and considering a single covariate.

3.1 Main result

Next, the asymptotic distribution of T_n is derived. The following assumptions on the stochastic nature of the observations, and on the nonparametric estimator of the regression function are needed:

(A1) The regression and the density functions m and f, respectively, are twice continuously differentiable.

(A2) The weight function w is continuously differentiable.

(A3) The marginal density f is continuous, bounded away from zero and $f(\mathbf{x}) > 0$ for all $\mathbf{x} \in D$.

(A4) For the correlation function ρ_n, there exist constants ρ_M and ρ_c such that $n \int |\rho_n(\mathbf{x})| d\mathbf{x} < \rho_M$ and $\lim_{n\to\infty} n \int \rho_n(\mathbf{x}) d\mathbf{x} = \rho_c$. For any sequence $\epsilon_n > 0$ satisfying $n^{1/d}\epsilon_n \to \infty$,

$$n \int_{\|\mathbf{x}\| \geq \epsilon_n} |\rho_n(\mathbf{x})| d\mathbf{x} \to 0 \quad \text{as} \quad n \to \infty.$$

(A5) For any i, j, k, l,

$$\text{Cov}(\varepsilon_i \varepsilon_j, \varepsilon_k \varepsilon_l) = \text{Cov}(\varepsilon_i, \varepsilon_k)\text{Cov}(\varepsilon_j, \varepsilon_l) + \text{Cov}(\varepsilon_i, \varepsilon_l)\text{Cov}(\varepsilon_j, \varepsilon_k).$$

(A6) It is assumed that errors are a geometrically strong mixing sequence with mean zero and $\mathbb{E}|\varepsilon(\mathbf{x})|^r < \infty$ for all $r > 4$.

(A7) The kernel K is a spherically symmetric density function, twice continuously differentiable and with compact support (for simplicity with a nonzero value only if $\|\mathbf{u}\| \leq 1$). Moreover, $\int \mathbf{u}\mathbf{u}' K(\mathbf{u}) d\mathbf{u} = \mu_2(K)\mathbf{I}_d$, where $\mu_2(K) \neq 0$ is scalar and $\mathbf{I}_d$ is the $d \times d$ identity matrix.

(A8) K is Lipschitz continuous. That is, there exists $\mathfrak{L} > 0$, such that

$$|K(\mathbf{X}_1) - K(\mathbf{X}_2)| \leq \mathfrak{L} \|\mathbf{X}_1 - \mathbf{X}_2\|, \quad \forall \mathbf{X}_1, \mathbf{X}_2 \in D.$$

(A9) The bandwidth matrix $\mathbf{H}$ is symmetric and positive definite, with $\mathbf{H} \to 0$ and $n|\mathbf{H}|\lambda_{\min}^2(\mathbf{H}) \to \infty$, when $n \to \infty$. The ratio $\lambda_{\max}(\mathbf{H})/\lambda_{\min}(\mathbf{H})$ is bounded above, where $\lambda_{\max}(\mathbf{H})$ and $\lambda_{\min}(\mathbf{H})$ are the maximum and minimum eigenvalues of $\mathbf{H}$, respectively.

As for the parametric estimator, just the assumption of being a $\sqrt{n}$-consistent estimator is required. This is guaranteed if the estimator $m_{\hat{\boldsymbol{\beta}}}$ described in Section 2.2 is employed in the statistic (6). Anyway, as pointed out in the previous section, a different parametric estimator of the regression function could be used in the test statistic (6) as long as this property was fulfilled.

Assumption (A4) implies that the correlation function depends on n, and the integral $\int |\rho_n(\mathbf{x})| d\mathbf{x}$ should vanish as $n \to \infty$. The vanishing speed should not be slower than $O(n^{-1})$. This assumption also implies that the integral of $|\rho_n(\mathbf{x})|$ is essentially dominated by the values of $\rho_n(\mathbf{x})$ near to the origin $\mathbf{0}$. Hence, the correlation is short-range and decreases as $n \to \infty$. Arguing somewhat loosely, this can be considered as a case of increasing-domain spatial asymptotics (see Cressie, 1993), since this setup can immediately be transformed to one in which the correlation function ρ_n is fixed with respect to the sample size, but the support D for $\mathbf{x}$ expands. The current setup with fixed domain D and shrinking ρ_n is more natural to consider when the primary purpose of the estimation is a fixed regression function m defined over a spatial domain, not the correlation function itself.

Two examples of commonly used correlation functions that satisfy the conditions of assumption (A4) are the exponential model

$$\rho_n(\mathbf{x}) = \exp(-an\|\mathbf{x}\|),$$

and the rational quadratic model

$$\rho_n(\mathbf{x}) = \frac{1}{1 + a(n\|\mathbf{x}\|)^2},$$

with $a > 0$ in both cases (see Cressie, 1993). In general, if $\rho_n(\mathbf{x}) = \rho(n^{1/d}\mathbf{x})$ and $\rho(\mathbf{x})$ is a fixed valid correlation function, which is continuous everywhere except at a finite number of points and absolutely integrable in $\mathbb{R}^d$, then it is easy to check that $\rho_n(\mathbf{x})$ satisfies assumption (A4).

Assumption (A5) is satisfied, for example, when the errors follow a Gaussian distribution. As for (A6), if $\mathcal{M}_a^b$ is the $\sigma-$field generated by $\{\xi(t) : a \le t \le b\}$, then $\{\xi(t) : t \in \mathbb{R}\}$ is geometrically strong mixing if the mixing coefficients verify

$$\alpha(\tau) = \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{M}_{-\infty}^0 \quad \text{and} \quad B \in \mathcal{M}_\tau^\infty\} = O(\zeta^\tau), \tag{8}$$

for some $0 < \zeta < 1$, when $\tau \to \infty$. This assumption is needed to apply the central limit theorem for reduced U-statistics under dependence given by Kim et al. (2013). Note that if a random variable is a real Gaussian process, the strong mixing coefficient and the correlation function are equivalent (Rozanov, 1967, p. 181). Therefore, hypotheses (A4)-(A6) could be satisfied by Gaussian error processes with exponential or rational quadratic (among others) correlation functions, having a decay rate larger than or equal to that indicated in (8).

In assumption (A9), $\mathbf{H} \to 0$ means that every entry of $\mathbf{H}$ goes to 0. Since $\mathbf{H}$ is symmetric and positive definite, $\mathbf{H} \to 0$ is equivalent to $\lambda_{\max}(\mathbf{H}) \to 0$.

$|\mathbf{H}|$ is a quantity of order $O(\lambda^d_{\max}(\mathbf{H}))$ because $|\mathbf{H}|$ is equal to the product of all eigenvalues of $\mathbf{H}$.

The following theorem shows the asymptotic distribution of the test statistic (6). A sketch of the proof is provided in the Appendix, while the detailed proof can be found in the Supplementary Material available in the Online Resource.

Theorem 1 *Under Assumptions* (A1)-(A9), *and if* $0 < V < \infty$, *it can be proved that*

$$V^{-1/2}(T_n - b_{0\mathbf{H}} - b_{1\mathbf{H}}) \to_{\mathcal{L}} N(0,1) \text{ as } n \to \infty,$$

where $\to_{\mathcal{L}}$ *denotes convergence in distribution, with*

$$b_{0\mathbf{H}} = |\mathbf{H}|^{-1/2}\sigma^2 K^{(2)}(\mathbf{0})\left[\int \frac{w(\mathbf{x})}{f(\mathbf{x})}d\mathbf{x} + \rho_c \int w(\mathbf{x})d\mathbf{x}\right],$$

$$b_{1\mathbf{H}} = \int (K_{\mathbf{H}} * g(\mathbf{x}))^2 w(\mathbf{x})d\mathbf{x},$$

and

$$V = 2\sigma^4 K^{(4)}(\mathbf{0})\left[\int \frac{w^2(\mathbf{x})}{f^2(\mathbf{x})}d\mathbf{x} + 2\rho_c \int \frac{w^2(\mathbf{x})}{f(\mathbf{x})}d\mathbf{x} + 4\rho_c^2 \int w^2(\mathbf{x})d\mathbf{x}\right],$$

where $K^{(j)}$ *denotes the* j*-times convolution product of* K *with itself.*

This result generalizes Theorem 2.1 of Alcalá et al. (1999) in the univariate case and with independent errors (corresponding to $\rho_c = 0$), considering the local polynomial regression estimator.

Remark 1 *The asymptotic distribution of the test statistic (6) can be also obtained under a geostatistical spatial trend model. In this scenario, model (1) can be viewed as an additive decomposition of the spatial process: the regression function* m *corresponds to the first-order moment of the process and captures the large-scale variability, whereas the error term collects the second-order structure, reflecting the small-scale variation. The covariates in this setting are given by the spatial locations (latitude and longitude), which are usually fixed in a geostatistical setting. In this case, considering assumptions (A1)-(A9), except the ones relative to* f *(given that we are under a fixed design scheme), and following similar steps to those employed in the proof of Theorem 1, but using Riemann approximations of sums by integrals, the asymptotic distribution of* T_n *is given by:*

$$V^{-1/2}(T_n - b_{0\mathbf{H}} - b_{1\mathbf{H}}) \to_{\mathcal{L}} N(0,1) \text{ as } n \to \infty,$$

with

$$b_{0\mathbf{H}} = |\mathbf{H}|^{-1/2}\sigma^2 K^{(2)}(\mathbf{0})\left[\int w(\mathbf{x})d\mathbf{x} + \rho_c \int w(\mathbf{x})d\mathbf{x}\right],$$

$$b_{1\mathbf{H}} = \int (K_{\mathbf{H}} * g(\mathbf{x}))^2 w(\mathbf{x})d\mathbf{x},$$

and

$$V = 2\sigma^4 K^{(4)}(\mathbf{0})\left[\int w^2(\mathbf{x})d\mathbf{x} + 2\rho_c \int w^2(\mathbf{x})d\mathbf{x} + 4\rho_c^2 \int w^2(\mathbf{x})d\mathbf{x}\right].$$

3.2 Calibration in practice

Once a suitable test statistic is available, a crucial task is the calibration of the critical value for a given level α, namely t_α. Usually, the determination of the critical value t_α, such that $\mathbb{P}_{H_0}(T_n \geq t_\alpha) = \alpha$ (denoting by $\mathbb{P}_{H_0}$ the probability under H_0), can be done by means of the asymptotic distribution of T_n. However, as noted in other nonparametric testing contexts, the asymptotic distribution obtained in Theorem 1 is often not sufficiently precise for constructing a practical test in small-to-medium sample size situation. Moreover, to use the asymptotic expression of T_n in practice, it is necessary to estimate some nuisance functions. The poor performance of the normal approximation for moderate sample sizes was observed in some simulation studies. A simple example, taking f and σ^2 as known, is included in the Supplementary Material.

Under these circumstances, calibration can be done by means of resampling procedures, such as bootstrap (see, for example, Francisco-Fernández et al., 2006). The bootstrap procedure considered (detailed below) extends to the case of spatially correlated data the parametric bootstrap discussed in Vilar-Fernández and González-Manteiga (1996). The specific steps are the following:

1. Obtain, using (5), the parametric regression estimator $\hat{\boldsymbol{\beta}}$.
2. Compute the estimated variance-covariance matrix of the errors, $\hat{\Sigma}$, using the residuals $\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n)'$, where $\hat{\varepsilon}_i = Z_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)$, $i = 1, \ldots, n$.
3. Find the matrix L, such that $\hat{\Sigma} = LL'$, using Cholesky decomposition.
4. Compute the "independent" variables, $\mathbf{e} = (e_1, \ldots, e_n)'$, given by $\mathbf{e} = L^{-1}\hat{\boldsymbol{\varepsilon}}$.
5. The previous independent variables are centered and an independent bootstrap sample of size n, denoted by $\mathbf{e}^* = (e_1^*, \ldots, e_n^*)$, is obtained.
6. Finally, the bootstrap errors $\boldsymbol{\varepsilon}^* = (\varepsilon_1^*, \ldots, \varepsilon_n^*)$ are $\boldsymbol{\varepsilon}^* = L\mathbf{e}^*$, and the bootstrap samples are $Z^*(\mathbf{X}_i) = m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i) + \varepsilon_i^*$.

Using the bootstrap sample $\{Z_i^*, i = 1, \ldots, n\}$, the bootstrap test statistic T_n^* is computed as in (6), by the weighted L_2-distance between the bootstrap versions of the smooth parametric fit (7) and the nonparametric estimator (3). Once the bootstrap statistic is obtained, the distribution of T_n^* can be approximated by Monte Carlo, and the $(1-\alpha)$ quantile t_α^* easily computed. Finally, the null hypothesis is rejected if $T_n > t_\alpha^*$.

4 Simulations

The finite sample performance of the proposed test, proceeding with a bootstrap calibration, is illustrated in this section with a simulation study. For this purpose, a linear parametric regression surface is chosen, $m_{\boldsymbol{\beta}}(X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$, being $\mathbf{X} = (X_1, X_2)$, and for different values of c the mean function

$$m(X_1, X_2) = 2 + X_1 + X_2 + cX_1^3 \tag{9}$$

is considered. Therefore, the parameter c controls whether the null ($c = 0$) or the alternative ($c \neq 0$) hypotheses are assumed. Values $c = 0$, 3, and 5 are considered in the study.

For each value of c, 500 samples of sizes $n = 225$ and 400 are generated on a bidimensional regular grid in the unit square, following model (1), with regression function (9) and random errors ε_i normally distributed with zero mean and isotropic exponential covariance function:

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = \sigma^2\{\exp(-\|\mathbf{X}_i - \mathbf{X}_j\|/a_e)\}, \tag{10}$$

where σ^2 is the variance and a_e is the practical range. Different degrees of spatial dependence were studied, considering values of $\sigma = 0.4$, 0.6, and 0.8, and $a_e = 0.1$ (weak correlation), $a_e = 0.2$ (medium correlation) and $a_e = 0.4$ (strong correlation). Note that no nugget effect is considered in this scenario.

To analyze the behavior of the test statistic given in (6) in the different scenarios, the bootstrap procedure described in Section 3.2 was applied, using $B = 500$ replications. The weight function used was taken constant with value 1. The nonparametric fit used for constructing (6) was obtained using the multivariate local linear estimator with a multiplicative triweight kernel, while the parametric fit was computed using the iterative least squares procedure, considering a linear model, these methods described in Sections 2.1 and 2.2, respectively. The bandwidth selection problem was addressed by using the same classical procedure as the one used in Härdle and Mammen (1993), Alcalá et al. (1999), or Opsomer and Francisco-Fernández (2010), among others. The test was run in a grid of several bandwidths to check how it is affected by the bandwidth choice. In order to simplify the calculations, the bandwidth matrix was restricted to a class of diagonal matrices with both equal elements (scalar matrices). To give a reasonable grid, the optimal bandwidth obtained by minimizing the mean average squared error (MASE) of the multivariate local linear estimator (see Francisco-Fernandez and Opsomer, 2005, p. 288) was calculated for each scenario. These bandwidths were in the interval $[0.6, 1]$, therefore, the bandwidth was taken as a diagonal matrix $\mathbf{H} = \text{diag}(h, h)$, and different values of h were chosen, $h = 0.6, 0.7, 0.8, 0.9, 1$.

Rejection proportions of the null hypothesis, for a significance level $\alpha = 0.05$, are displayed in Table 1, where it can be observed that the test has a reasonable behavior. If $c = 0$ (null hypothesis), the rejection proportions are similar to the theoretical level, although these proportions are affected by the value of h. In fact, in most of the cases, the rejection proportions are smaller when the bandwidth value is larger. As expected, considering a larger sample size, the bandwidth value should be smaller. For alternative assumptions ($c = 3$ and $c = 5$), a decreasing power of the test when the values of h increase is observed. For all the scenarios, the power of the test becomes

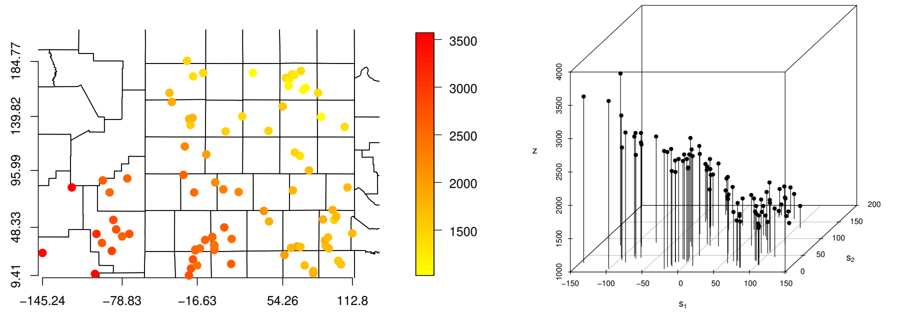

Fig. 1 Locations with the levels of piezometric-head for the Wolfcamp Aquifer (left) and its own 3-dimensional representation (right).

larger as the value of c increases. As expected, large values of the variance σ^2 lead to a decrease in power. Regarding the effect of the range a_e, when this parameter is larger, the power of the test increases, which justifies the correct performance of the bootstrap procedure for dependent data considered. It can be also noticed that, for large values of a_e, the bandwidth values providing an effective calibration of the test are also large.

Additional simulation studies with other regression functions, selecting bandwidth matrices with different values in the main diagonal, including a nugget effect and considering random designs were also performed, obtaining similar results to those shown in Table 1. These experiments are reported in the Supplementary Material.

5 Illustration with real data

In order to illustrate the performance in practice of the test statistic T_n, given in (6), the Wolfcamp aquifer dataset briefly mentioned in the Introduction is considered. These data were reported and geostatistically analyzed in Harper and Furr (1986) and Cressie (1993), and are available in the `R` package `npsp` (Fernández-Casal, 2016).

The Deaf Smith County (Texas, bordering New Mexico) was selected as an alternate site for a possible nuclear waste disposal repository in the 1980s. This site was later dropped on grounds of contamination of the aquifer, the source of much of the water supply for west Texas. In a study conducted by the U.S. Department of Energy, piezometric-head levels were obtained irregularly at 85 locations, shown in the left panel of Figure 1, by drilling a narrow pipe through the aquifer (see Harper and Furr, 1986). With higher values generally in the lower left (southwest) and lower values in the upper right (northwest), the groundwater gradient would cause water to flow in a northeasterly direction from the repository in Deaf Smith County toward Amarillo in lower Potter county.

Table 1 Rejection proportions of the null hypothesis for $\alpha = 0.05$.

				h				
σ	a_e	c	n	0.6	0.7	0.8	0.9	1
0.4	0.1	0	225	0.092	0.068	0.050	0.038	0.024
			400	0.050	0.036	0.024	0.022	0.020
0.4	0.1	3	225	0.522	0.480	0.458	0.446	0.458
			400	0.438	0.396	0.360	0.360	0.368
0.4	0.1	5	225	0.988	0.984	0.978	0.980	0.984
			400	1.000	1.000	1.000	1.000	1.000
0.4	0.2	0	225	0.082	0.062	0.048	0.032	0.022
			400	0.078	0.050	0.032	0.028	0.014
0.4	0.2	3	225	0.902	0.876	0.854	0.840	0.834
			400	0.896	0.870	0.832	0.818	0.806
0.4	0.2	5	225	1.000	1.000	1.000	1.000	1.000
			400	1.000	1.000	1.000	1.000	1.000
0.4	0.4	0	225	0.162	0.126	0.084	0.074	0.068
			400	0.164	0.126	0.098	0.076	0.058
0.4	0.4	3	225	0.978	0.976	0.974	0.970	0.970
			400	0.990	0.990	0.988	0.986	0.986
0.4	0.4	5	225	1.000	1.000	1.000	1.000	1.000
			400	1.000	1.000	1.000	1.000	1.000
0.6	0.1	0	225	0.090	0.068	0.054	0.038	0.026
			400	0.050	0.036	0.022	0.022	0.020
0.6	0.1	3	225	0.096	0.084	0.062	0.056	0.066
			400	0.082	0.058	0.046	0.034	0.036
0.6	0.1	5	225	0.684	0.652	0.624	0.602	0.608
			400	0.630	0.576	0.538	0.532	0.536
0.6	0.2	0	225	0.082	0.060	0.046	0.034	0.024
			400	0.074	0.050	0.032	0.028	0.014
0.6	0.2	3	225	0.492	0.430	0.370	0.332	0.322
			400	0.466	0.408	0.362	0.334	0.330
0.6	0.2	5	225	0.964	0.958	0.942	0.930	0.920
			400	0.962	0.948	0.930	0.916	0.912
0.6	0.4	0	225	0.158	0.126	0.084	0.074	0.068
			400	0.164	0.126	0.100	0.076	0.058
0.6	0.4	3	225	0.766	0.742	0.716	0.694	0.684
			400	0.818	0.784	0.744	0.714	0.704
0.6	0.4	5	225	0.998	0.998	0.998	0.998	0.998
			400	0.996	0.996	0.994	0.994	0.994
0.8	0.1	0	225	0.088	0.066	0.052	0.038	0.026
			400	0.050	0.036	0.022	0.022	0.020
0.8	0.1	3	225	0.080	0.052	0.036	0.030	0.026
			400	0.046	0.018	0.008	0.006	0.006
0.8	0.1	5	225	0.282	0.240	0.204	0.196	0.204
			400	0.190	0.158	0.128	0.120	0.126
0.8	0.2	0	225	0.082	0.060	0.046	0.032	0.024
			400	0.076	0.050	0.032	0.028	0.014
0.8	0.2	3	225	0.282	0.212	0.174	0.142	0.146
			400	0.256	0.202	0.164	0.144	0.144
0.8	0.2	5	225	0.716	0.654	0.614	0.588	0.574
			400	0.704	0.654	0.628	0.600	0.572
0.8	0.4	0	225	0.158	0.124	0.084	0.074	0.068
			400	0.164	0.126	0.100	0.074	0.058
0.8	0.4	3	225	0.556	0.496	0.458	0.434	0.426
			400	0.580	0.532	0.484	0.450	0.430
0.8	0.4	5	225	0.928	0.920	0.906	0.888	0.874
			400	0.952	0.940	0.930	0.920	0.904

Figure 1 (right panel) displays the 3-dimensional scatterplot of the piezometric heads levels (feet above sea level) against the coordinates (miles, from a reference point). This plot shows a clear downwards trend from south-west to north-east. Cressie (1993) used the median polish approach to model this trend, whereas Harper and Furr (1986) considered a linear trend surface, that is, a linear regression model on latitude and longitude. In order to check if a linear model is plausible, the test T_n, using the bootstrap procedure described in Section 3.2 with $B = 1000$ replications, was applied considering a linear parametric model, $m_{\boldsymbol{\beta}}(X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$, as the null hypothesis, being X_1 and X_2 the spatial coordinates of the points where the process is observed. It should be noted that the (nonparametric) detrended data were also tested for isotropy and stationarity, following the proposals by Bowman and Crujeiras (2013), obtaining p-values of 0.838 for isotropy and 0.031 for stationarity.

To apply the test (6), the parametric fit was carried out using the iterative least squares estimator described in Section 2.2, assuming a linear regression model. After analyzing the initial residuals obtained by least squares regression, a spherical correlation model (as it was suggested by Harper and Furr, 1986) was considered to estimate the variance-covariance matrix of the errors, needed to obtain a feasible estimate of $\boldsymbol{\beta}$. As for the nonparametric fit in (6), the local linear estimator (3) with a multiplicative triweight kernel was considered. The bandwidth was taken as a diagonal matrix $\mathbf{H} = \text{diag}(h_1, h_2)$, being the values of h_1 and h_2 different. Note that the corrected generalized cross-validation bandwidth (Francisco-Fernandez and Opsomer, 2005; Francisco-Fernández et al., 2012) is $\mathbf{H} = \text{diag}(403.19, 226.20)$.

Figure 2 shows the smooth version of the parametric (left panel) and the nonparametric (right panel) regression estimators using the corrected generalized cross-validation bandwidth for the level of piezometric-head in the area of study. These regression surfaces are compared in the proposed test statistic. Figure 3 shows the p-values of the test using the so-called significance trace (Bowman and Azzalini, 1997), that is, the proportions of empirical rejections for different bandwidths. Taking into account this plot, there are no evidences against a linear spatial regression. Note that smaller bandwidths than those considered should not be taken to avoid boundary problems.

6 Discussion

A goodness-of-fit test for a parametric regression model with correlated errors is presented in this work, based on a L_2-distance between a parametric and a nonparametric fits. A least squares procedure has been considered as a parametric approach, given its efficiency, but other methods such as maximum likelihood methods, could be also used, as long as a $\sqrt{n}-$consistency property is satisfied. In this case, it should be noted that both the regression function and the dependence structure of the errors are jointly estimated, but usually restricted to a (multivariate) Gaussian distribution of the process realization.

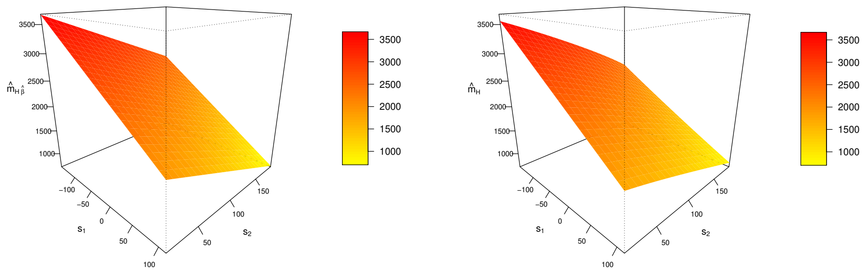

Fig. 2 Smooth version of the parametric fit (left) and nonparametric estimator of the regression (right) using the corrected generalized cross-validation bandwidth for the Wolfcamp Aquifer.

In both cases (least squares and maximum likelihood), a parametric form for the correlation is considered. Without being the target function and viewing spatial correlation as a nuisance (that should certainly be accounted for, but it is not of primary interest), it is expected that the proposed goodness-of-fit test has a good performance even when the correlation is misspecified as long as it can be reasonably well approximated. Testing approaches as those proposed in Maglione and Diblasi (2004) can be useful for this task. Regarding the nonparametric counterpart in the test statistic, other kernel estimators such as Priestley-Chao or Nadaraya-Watson estimators could be used.

Asymptotic results, under the null and under local alternatives, support the proposal but due to the slow convergence to the limit distribution, a bootstrap procedure is presented. Simulation results confirm that the bootstrap algorithm works, facilitating the practical application of the test, with no other competitor (up to our knowledge). It may be argued that this simulation study was limited to bidimensional linear regression models, but it could be extended to any parametric family. It should be noted (Cressie, 1993; Diggle and Ribeiro, 2007) that, in the geostatistical context, simple parametric models are usually preferred in order to preserve interpretability. If one would be interested in a more sophisticated regression structure, then a nonparametric fit could provide an appealing alternative. In any case, the bandwidth matrix needed to apply (6) can be selected by cross-validation but recall that this bandwidth is not necessary a good one for testing. With this purpose, it is advisable to explore a range of bandwidths, taking a data-driven one as a reference.

Although a homoscedastic regression model has been considered in this paper, under suitable assumptions, the asymptotic results of the test statistic could be also derived for certain heteroscedastic regression models. In such a context, the bootstrap method to calibrate the test, described in Section 3.2, could be also modified, using an appropriate route to estimate the dependence of the model. To do this, the nonparametric approach described by Fernández-Casal et al. (2017) could be used. Note that in that case, due to heteroscedasticity, the use of a wild bootstrap procedure in the resampling pro-

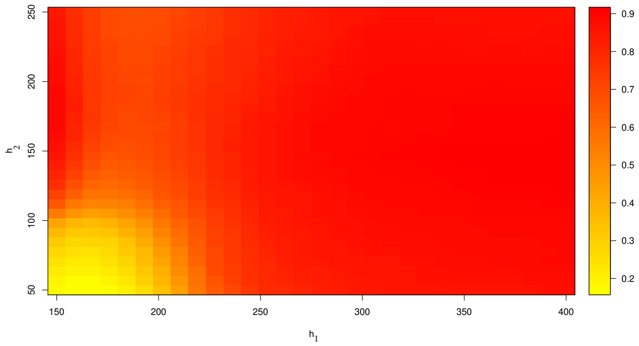

Fig. 3 Significance trace of the test for $\alpha = 0.05$ for the Wolfcamp aquifer dataset.

cess could be more convenient. The design of this type of resampling approach in this context is, indeed, an interesting issue for a future research.

The procedures used in the simulation study as well as in the illustration with real data were implemented in the statistical environment `R` (R Development Core Team, 2019), using functions included in the `geoR` and `npsp` packages (Ribeiro and Diggle, 2016; Fernández-Casal, 2016) to estimate the variogram and the spatial regression functions.

Acknowledgements The authors acknowledge the support from the Xunta de Galicia and the European Union (European Social Fund - ESF). This research has been partially supported by MINECO grants MTM2014-52876-R, MTM2016-76969-P and MTM2017-82724-R, and by the Xunta de Galicia (Grupos de Referencia Competitiva ED431C-2016-015 and ED431C- 2017-38, and Centro Singular de Investigación de Galicia ED431G/01), all of them through the ERDF. We also thank two reviewers and the Associate Editor for their helpful comments and suggestions that significantly improved this article.

Appendix A. Proof of Theorem 1

The test statistic (6) can be written as

$$\begin{aligned}
T_n &= n|\mathbf{H}|^{1/2} \int (\hat{m}^{LL}_{\mathbf{H}}(\mathbf{x}) - \hat{m}^{LL}_{\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x}))^2 w(\mathbf{x}) d\mathbf{x} \\
&= n|\mathbf{H}|^{1/2} \int \left[e_1' \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) & \frac{1}{n}\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})' \\ \frac{1}{n}\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x}) & \frac{1}{n}\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})' \end{pmatrix}^{-1} \right. \\
&\quad \left. \cdot \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(Z_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)) \\ \frac{1}{n}\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})(Z_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)) \end{pmatrix} \right]^2 w(\mathbf{x}) d\mathbf{x}.
\end{aligned}$$

Taking into account that, for every $\eta > 0$, $\hat{f}_{\mathbf{H}}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) = f(\mathbf{x}) + O_p(n^{-2/(4+d)+\eta})$ uniformly in $\mathbf{x}$ (see Härdle and Mammen, 1993), and according to Liu (2001), it follows that

$$
\begin{aligned}
T_n &= n|\mathbf{H}|^{1/2} \int \left[\frac{1}{nf(\mathbf{x})} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(Z_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)) \right. \\
&\quad \left. - \nabla f(\mathbf{x}) \frac{1}{nf^2(\mathbf{x})} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})(Z_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)) \right]^2 w(\mathbf{x}) d\mathbf{x} + O_p(n^{-2/(4+d)+\eta}) \\
&= T_{n1} + T_{n2} + 2T_{n12} + O_p(n^{-2/(4+d)+\eta}), \qquad (11)
\end{aligned}
$$

where $\nabla f(\mathbf{x})$ denotes the $d \times 1$ vector of first-order partial derivatives of f, and

$$
\begin{aligned}
T_{n1} &= n|\mathbf{H}|^{1/2} \int \left[\frac{1}{nf(\mathbf{x})} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(Z_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)) \right]^2 w(\mathbf{x}) d\mathbf{x}, \\
T_{n2} &= n|\mathbf{H}|^{1/2} \int \left[\nabla f(\mathbf{x}) \frac{1}{nf^2(\mathbf{x})} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})(Z_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)) \right]^2 w(\mathbf{x}) d\mathbf{x},
\end{aligned}
$$

denoting by T_{n12} the integral of the cross product. Regarding T_{n1}, taking into account that the regression models considered are of the form $m(\mathbf{x}) = m_{\boldsymbol{\beta}_0}(\mathbf{x}) + n^{-1/2}|\mathbf{H}|^{-1/4} g(\mathbf{x})$, one gets that

$$
\begin{aligned}
T_{n1} &= n|\mathbf{H}|^{1/2} \int \left[\frac{1}{nf(\mathbf{x})} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(m_{\boldsymbol{\beta}_0}(\mathbf{X}_i) + n^{-1/2}|\mathbf{H}|^{-1/4} g(\mathbf{X}_i) \right. \\
&\quad \left. + \varepsilon_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)) \right]^2 w(\mathbf{x}) d\mathbf{x} \\
&= n|\mathbf{H}|^{1/2} \int \frac{1}{f^2(\mathbf{x})} (I_1(\mathbf{x}) + I_2(\mathbf{x}) + I_3(\mathbf{x}))^2 w(\mathbf{x}) d\mathbf{x},
\end{aligned}
$$

where

$$
\begin{aligned}
I_1(\mathbf{x}) &= \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}}\left(\mathbf{X}_i - \mathbf{x}\right) \left(m_{\boldsymbol{\beta}_0}(\mathbf{X}_i) - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)\right), \\
I_2(\mathbf{x}) &= \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}}\left(\mathbf{X}_i - \mathbf{x}\right) n^{-1/2}|\mathbf{H}|^{-1/4} g(\mathbf{X}_i), \\
I_3(\mathbf{x}) &= \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}}\left(\mathbf{X}_i - \mathbf{x}\right) \varepsilon_i.
\end{aligned}
$$

Under assumptions (A1)–(A3) and (A7), and given that the difference $m_{\hat{\boldsymbol{\beta}}}(\mathbf{x}) - m_{\boldsymbol{\beta}_0}(\mathbf{x}) = O_p(n^{-1/2})$ uniformly in $\mathbf{x}$, it is obtained that

$$
n|\mathbf{H}|^{1/2} \int \frac{1}{f^2(\mathbf{x})} I_1^2(\mathbf{x}) w(\mathbf{x}) d\mathbf{x} = O_p(|\mathbf{H}|^{1/2}). \qquad (12)
$$

As for the term $I_2(\mathbf{x})$, taking into account Lemma 1 (available in the Online Supplementary Material), by straightforward calculations it follows that

$$n|\mathbf{H}|^{1/2}\int \frac{1}{f^2(\mathbf{x})}I_2^2(\mathbf{x})w(\mathbf{x})d\mathbf{x}$$
$$= \int (K_{\mathbf{H}} * g)^2(\mathbf{x})w(\mathbf{x})d\mathbf{x}\cdot\{1+o_p(1)\}. \tag{13}$$

Note that the leading term of (13) is the term $b_{1\mathbf{H}}$ in Theorem 1. Finally, $I_3(\mathbf{x})$ (associated with the error component) can be decomposed as

$$n|\mathbf{H}|^{1/2}\int \frac{1}{f^2(\mathbf{x})}I_3^2(\mathbf{x})w(\mathbf{x})d\mathbf{x} = n|\mathbf{H}|^{1/2}\int \frac{1}{f^2(\mathbf{x})n^2}\sum_{i=1}^{n} K_{\mathbf{H}}^2\left(\mathbf{X}_i-\mathbf{x}\right)\varepsilon^2(\mathbf{X}_i)w(\mathbf{x})d\mathbf{x}$$
$$+n|\mathbf{H}|^{1/2}\int \frac{1}{f^2(\mathbf{x})n^2}\sum_{i\neq j} K_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{x}\right)K_{\mathbf{H}}\left(\mathbf{X}_j-\mathbf{x}\right)\varepsilon_i\varepsilon_j w(\mathbf{x})d\mathbf{x}$$
$$= I_{31}+I_{32}.$$

For the first term, one gets that

$$\mathbb{E}(I_{31}) = \mathbb{E}\left[\sigma^2 n|\mathbf{H}|^{1/2}\int \frac{1}{f^2(\mathbf{x})}\frac{1}{n^2}\sum_{i=1}^{n} K_{\mathbf{H}}^2\left(\mathbf{X}_i-\mathbf{x}\right)w(\mathbf{x})d\mathbf{x}\right]$$
$$= \sigma^2|\mathbf{H}|^{-1/2}K^{(2)}(\mathbf{0})\int \frac{w(\mathbf{x})}{f(\mathbf{x})}d\mathbf{x}\cdot\{1+o(1)\}.$$

Similarly, it is obtained that $\mathrm{Var}(I_{31}) = O_p(n^{-1}|\mathbf{H}|^{-1})$, and, therefore,

$$I_{31} = \sigma^2|\mathbf{H}|^{-1/2}K^{(2)}(\mathbf{0})\int \frac{w(\mathbf{x})}{f(\mathbf{x})}d\mathbf{x}\cdot\{1+o_p(1)\}. \tag{14}$$

The leading term of (14) corresponds to the first term of $b_{0\mathbf{H}}$ in Theorem 1. For the term I_{32}, let

$$\kappa_{ij} = n|\mathbf{H}|^{1/2}\int \frac{1}{f^2(\mathbf{x})}\frac{1}{n^2}K_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{x}\right)K_{\mathbf{H}}\left(\mathbf{X}_j-\mathbf{x}\right)\varepsilon_i\varepsilon_j w(\mathbf{x})d\mathbf{x},$$

thus,

$$I_{32} = \sum_{i\neq j}\kappa_{ij},$$

and this can be seen as a U-statistic with degenerate kernel. To obtain the asymptotic normality of I_{32}, considering assumption (A6), Theorem 2 given in Kim et al. (2013) will be applied. For this term, under assumptions (A4), (A7), (A8) and (A9), and according to Liu (2001), one gets that

$$\mathbb{E}(I_{32}) = \frac{n-1}{n}|\mathbf{H}|^{-1/2}\sigma^2\int\Bigg(n|\mathbf{H}|\iint K\left(\mathbf{p}\right)K\left(\mathbf{q}\right)\rho_n(\mathbf{H}(\mathbf{p}-\mathbf{q}))d\mathbf{p}d\mathbf{q}$$
$$+\, o(1)\Bigg)w(\mathbf{x})d\mathbf{x}$$
$$= |\mathbf{H}|^{-1/2}\sigma^2K^{(2)}(0)\rho_c\int w(\mathbf{x})d\mathbf{x}\cdot\{1+o(1)\}, \tag{15}$$

corresponding to the second term of $b_{0\mathbf{H}}$ in Theorem 1.

Similarly, it can be shown that the leading term of the variance of I_{32} is given by:

$$V = 2\sigma^4 K^{(4)}(0)\left[\int \frac{w^2(\mathbf{x})}{f^2(\mathbf{x})}d\mathbf{x} + 2\rho_c \int \frac{w^2(\mathbf{x})}{f(\mathbf{x})}d\mathbf{x} + 4\rho_c^2 \int w^2(\mathbf{x})d\mathbf{x}\right]. \quad (16)$$

Therefore, using the central limit theorem for degenerate reduced U-statistics under α-mixing conditions, given in Kim et al. (2013), it is obtained that the term I_{32} converges, in distribution, to a normal distribution with mean the leading term of (15) and variance (16).

On the other hand, in virtue of the Cauchy-Schwarz inequality, the cross terms in T_{n1} resulting from the products of $I_1(\mathbf{x})$, $I_2(\mathbf{x})$ and $I_3(\mathbf{x})$ are all of smaller order. Therefore, combining the results given in (12), (13) and (14), and the asymptotic normality of I_{32} (with bias the leading term of (15) and variance (16)), one gets

$$V^{-1/2}(T_{n1} - b_{0\mathbf{H}} - b_{1\mathbf{H}}) \to_{\mathcal{L}} N(0,1) \text{ as } n \to \infty, \quad (17)$$

where

$$b_{0\mathbf{H}} = |\mathbf{H}|^{-1/2}\sigma^2 K^{(2)}(\mathbf{0})\left[\int \frac{w(\mathbf{x})}{f(\mathbf{x})}d\mathbf{x} + \rho_c \int w(\mathbf{x})d\mathbf{x}\right],$$

$$b_{1\mathbf{H}} = \int (K_{\mathbf{H}} * g)^2(\mathbf{x})w(\mathbf{x})d\mathbf{x},$$

and

$$V = 2\sigma^4 K^{(4)}(0)\left[\int \frac{w^2(\mathbf{x})}{f^2(\mathbf{x})}d\mathbf{x} + 2\rho_c \int \frac{w^2(\mathbf{x})}{f(\mathbf{x})}d\mathbf{x} + 4\rho_c^2 \int w^2(\mathbf{x})d\mathbf{x}\right].$$

The term T_{n2} in T_n is of smaller order than T_{n1} (specifically, $T_{n2} = O_p(\mathrm{tr}(\mathbf{H}^2)T_{n1})$), and by the Cauchy-Schwarz inequality, the cross term T_{n12} is of smaller order as well. Therefore, from (11), it follows that

$$T_n = T_{n1} + O_p(tr(\mathbf{H}^2)) + O_p(n^{-2/(4+d)+\eta}).$$

Taking into account (17), it follows that

$$V^{-1/2}(T_n - b_{0\mathbf{H}} - b_{1\mathbf{H}}) \to_{\mathcal{L}} N(0,1) \text{ as } n \to \infty.$$

with $b_{0\mathbf{H}}$, $b_{1\mathbf{H}}$ and V given above.

References

Alcalá J, Cristóbal J, González-Manteiga W (1999) Goodness-of-fit test for linear models based on local polynomials. Statist Probab Lett 42:39–46

Azzalini A, Bowman AW, Härdle W (1989) On the use of nonparametric regression for model checking. Biometrika 76:1–11

Biedermann S, Dette H (2000) Testing linearity of regression models with dependent errors by kernel based methods. Test 9:417–438

Bowman AW, Azzalini A (1997) Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations, vol 18. OUP Oxford

Bowman AW, Crujeiras RM (2013) Inference for variograms. Computational Statistics & Data Analysis 66:19–31

Cressie N (1985) Fitting variogram models by weighted least squares. J Int Ass Math Geol 17:563–586

Cressie NA (1993) Statistics for spatial data. Wiley, New York

Crujeiras RM, Van Keilegon I (2010) Least squares estimation of nonlinear spatial trends. Comput Stat Data Anal 54:452–465

Diblasi A, Bowman A (2001) On the use of the variogram in checking for independence in spatial data. Biometrics 57:211–218

Diggle P, Ribeiro PJ (2007) Model-based geostatistics. Springer, New York

Eubank RL, Spiegelman CH (1990) Testing the goodness of fit of a linear model via nonparametric regression techniques. J Am Stat Assoc 85:387–392

Eubank RL, Li CS, Wang S (2005) Testing lack-of-fit of parametric regression models using nonparametric regression techniques. Stat Sin 15:135–152

Fan J, Gijbels I (1996) Local polynomial modelling and its applications. Chapman and Hall, London

Fernández-Casal R (2016) `npsp`: Nonparametric spatial (geo)statistics. URL `http://cran.r-project.org/package=npsp`, R package version 0.5-3

Fernández-Casal R, Castillo-Páez S, García-Soidán P (2017) Nonparametric estimation of the small-scale variability of heteroscedastic spatial processes. Spat Stat 22:358–370

Francisco-Fernandez M, Opsomer JD (2005) Smoothing parameter selection methods for nonparametric regression with spatially correlated errors. Can J Stat-Rev Can Stat 33:279–295

Francisco-Fernández M, Jurado-Expósito M, Opsomer J, López-Granados F (2006) A nonparametric analysis of the spatial distribution of *Convolvulus arvensis* in wheat-sunflower rotations. Environmetrics 17:849–860

Francisco-Fernández M, Quintela-del Río A, Fernández-Casal R (2012) Nonparametric methods for spatial regression. An application to seismic events. Environmetrics 23(1):85–93

González-Manteiga W, Crujeiras RM (2013) An updated review of Goodness-of-Fit tests for regression models. Test 22:361–411

González-Manteiga W, Vilar-Fernández J (1995) Testing linear regression models using non-parametric regression estimators when errors are non-independent. Comput Stat Data Anal 20:521–541

Hallin M, Lu Z, Tran LT (2004) Local linear spatial regression. Ann Stat 32:2469–2500

Härdle W, Mammen E (1993) Comparing nonparametric versus parametric regression fits. Ann Stat 21:1926–1947

Harper WV, Furr JM (1986) Geostatistical analysis of potentiometric data in Wolfcamp aquifer of the Palo Duro Basin, Texas. Tech. rep., Battelle Memorial Inst.

Kim TY, Ha J, Hwang SY, Park C, Luo ZM (2013) Central limit theorems for reduced U-statistics under dependence and their usefulness. Aust N Z J Stat 55:387–399

Li CS (2005) Using local linear kernel smoothers to test the lack of fit of nonlinear regression models. Stat Methodol 2:267–284

Liu XH (2001) Kernel smoothing for spatially correlated data. PhD thesis, Department of Statistics, Iowa State University

Maglione D, Diblasi A (2004) Exploring a valid model for the variogram of an isotropic spatial process. Stoch Environ Res Risk Assess 18:366–376

Nadaraya EA (1964) On estimating regression. Theory Probab Appl 9:141–142

Opsomer J, Francisco-Fernández M (2010) Finding local departures from a parametric model using nonparametric regression. Stat Pap 51:69–84

Park C, Kim TY, Ha J, Luo ZM, Hwang SY (2015) Using a bimodal kernel for a nonparametric regression specification test. Stat Sin 25:1145–1161

R Development Core Team (2019) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL `http://www.R-project.org`

Ribeiro PJ, Diggle PJ (2016) geoR: Analysis of Geostatistical Data. URL `https://cran.r-project.org/package=geoR`, R package version 1.7-5.2

Rozanov YA (1967) Stationary random processes. Holden Day, Oakland, CA

Ruppert D, Wand MP (1994) Multivariate locally weighted least squares regression. Ann Stat pp 1346–1370

Vilar-Fernández J, González-Manteiga W (1996) Bootstrap test of goodness of fit to a linear model when errors are correlated. Commun Stat-Theory Methods 25:2925–2953

Watson GS (1964) Smooth regression analysis. Sankhya 26:359–372

Weihrather G (1993) Testing a linear regression model against nonparametric alternatives. Metrika 40:367–379

Supplementary Material for "A goodness-of-fit test for regression models with spatially correlated errors"

Andrea Meilán-Vila
Universidade da Coruña*

Jean D. Opsomer
Westat†

Mario Francisco-Fernández
Universidade da Coruña*

Rosa M. Crujeiras
Universidade de Santiago de Compostela‡

This supplementary material for "A goodness-of-fit test for regression models with spatially correlated errors" provides a detailed proof of the main theorem, jointly with the required auxiliary lemmas. Some additional simulation results, illustrating the asymptotic distribution of the test, the use of other type of bandwidths and the performance of the test in different settings (random design and nugget effect in the dependence structure) are included in this document.

1 Theoretical results

Theorem 1 in the paper establishes the asymptotic distribution of

$$T_n = n|\mathbf{H}|^{1/2} \int_D (\hat{m}^{LL}_{\mathbf{H}}(\mathbf{x}) - \hat{m}^{LL}_{\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x}))^2 w(\mathbf{x}) d\mathbf{x}, \tag{1}$$

defined in Equation (6), under the null and under local alternatives. Denoting by $\boldsymbol{\beta}_0$ the "true" parameter under the null hypothesis, the types of model deviations that can be captured by this test are of the form $m(\mathbf{x}) = m_{\boldsymbol{\beta}_0}(\mathbf{x}) + c_n g(\mathbf{x})$, where c_n is a sequence, such that $c_n \to 0$ and g is a deterministic function collecting the deviation direction for the alternative hypothesis. In Theorem 1, the asymptotic distribution of the test statistic (1) is stablished under the null hypothesis, and also under local alternatives converging to the null hypothesis at a certain rate controlled by c_n. Specifically, it is assumed that the function g is bounded uniformly in $\mathbf{x}$ and n, and $c_n = n^{-1/2}|\mathbf{H}|^{-1/4}$. In particular, this setting includes the null hypothesis corresponding to $g(\mathbf{x}) = 0$. The assumptions required for that result are the following:

(A1) The regression and the density functions m and f, respectively, are twice continuously differentiable.

*Research group MODES, Department of Mathematics, Faculty of Computer Science, Universidade da Coruña, Campus de Elviña s/n, 15071, A Coruña, Spain

†Westat 1600 Research Boulevard Rockville, MD, 20850, Maryland, USA

‡Department of Statistics, Mathematical Analysis and Optimization, Faculty of Mathematics, Universidade de Santiago de Compostela, Rúa Lope Gómez de Marzoa s/n, 15782, Santiago de Compostela, Spain

(A2) The weight function w is continuously differentiable.

(A3) The marginal density f is continuous, bounded away from zero and $f(\mathbf{x}) > 0$ for all $\mathbf{x} \in D$.

(A4) For the correlation function ρ_n, there exist constants ρ_M and ρ_c such that $n\int|\rho_n(\mathbf{x})|d\mathbf{x} < \rho_M$ and $\lim_{n\to\infty} n\int\rho_n(\mathbf{x})d\mathbf{x} = \rho_c$. For any sequence $\epsilon_n > 0$ satisfying $n^{1/d}\epsilon_n \to \infty$,

$$n\int_{\|\mathbf{x}\|\geq\epsilon_n} |\rho_n(\mathbf{x})|d\mathbf{x} \to 0 \quad \text{as} \quad n \to \infty.$$

(A5) For any i, j, k, l,

$$\text{Cov}(\varepsilon_i\varepsilon_j, \varepsilon_k\varepsilon_l) = \text{Cov}(\varepsilon_i, \varepsilon_k)\text{Cov}(\varepsilon_j, \varepsilon_l) + \text{Cov}(\varepsilon_i, \varepsilon_l)\text{Cov}(\varepsilon_j, \varepsilon_k).$$

(A6) It is assumed that the errors are a geometrically strong mixing sequence with mean zero and $\mathbb{E}|\varepsilon|^r < \infty$ for all $r > 4$.

(A7) The kernel K is a spherically symmetric density function, twice continuously differentiable and with compact support (for simplicity with a nonzero value only if $\|\mathbf{u}\| \leq 1$). Moreover, $\int \mathbf{u}\mathbf{u}'K(\mathbf{u})d\mathbf{u} = \mu_2(K)\mathbf{I}_d$, where $\mu_2(K) \neq 0$ is scalar and $\mathbf{I}_d$ is the $d \times d$ identity matrix.

(A8) K is Lipschitz continuous. That is, there exists an $\mathfrak{L} > 0$, such that

$$|K(\mathbf{X}_1) - K(\mathbf{X}_2)| \leq \mathfrak{L}\|\mathbf{X}_1 - \mathbf{X}_2\|, \quad \forall \mathbf{X}_1, \mathbf{X}_2 \in D.$$

(A9) The $d\times d$ bandwidth matrix $\mathbf{H}$ is symmetric and positive definite, with $\mathbf{H} \to 0$ and $n|\mathbf{H}|\lambda^2_{\min}(\mathbf{H}) \to \infty$, when $n \to \infty$. The ratio $\lambda_{\max}(\mathbf{H})/\lambda_{\min}(\mathbf{H})$ is bounded above, where $\lambda_{\max}(\mathbf{H})$ and $\lambda_{\min}(\mathbf{H})$ are the maximum and minimum eigenvalues of $\mathbf{H}$, respectively.

In what follows, $\mathbf{1}_d$ and $\mathbf{1}_{d\times d}$ are used to denote the $d \times 1$ vector and the $d \times d$ matrix with all entries equal to 1, respectively. Moreover, if $\mathbf{U}_n$ is a random matrix, then $O_p(\mathbf{U}_n)$ and $o_p(\mathbf{U}_n)$ are to be taken componentwise.

Before deriving the proof of Theorem 1, some auxiliary lemmas are required.

Lemma 1. *Let*

$$W_{1n}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})\, g(\mathbf{X}_i),$$

where g is a bounded function uniformly at $\mathbf{x}$. For any $\mathbf{x} \in D$, under assumptions (A1),(A3), (A7) and (A9), one gets that

$$W_{1n}(\mathbf{x}) = \int K(\mathbf{p})\, g(\mathbf{x} + \mathbf{H}\mathbf{p})f(\mathbf{x} + \mathbf{H}\mathbf{p})d\mathbf{p} + o_p(1).$$

Proof of Lemma 1. For any $\mathbf{x} \in D$, under assumptions (A1),(A3), (A7) and (A9), it follows that

$$\begin{aligned} E(W_{1n}(\mathbf{x})) &= \int K_{\mathbf{H}}(\mathbf{u} - \mathbf{x})\, g(\mathbf{u})f(\mathbf{u})d\mathbf{u} \\ &= \int K(\mathbf{p})\, g(\mathbf{x} + \mathbf{H}\mathbf{p})f(\mathbf{x} + \mathbf{H}\mathbf{p})d\mathbf{p} \end{aligned}$$

and

$$
\begin{aligned}
\mathrm{Var}(W_{1n}(\mathbf{x})) &\leq \frac{1}{n}\int K_{\mathbf{H}}^{2}\left(\mathbf{u}-\mathbf{x}\right)g^{2}(\mathbf{u})f(\mathbf{u})d\mathbf{u} \\
&= \frac{1}{n|\mathbf{H}|}\int K^{2}\left(\mathbf{p}\right)g^{2}(\mathbf{x}+\mathbf{H}\mathbf{p})f(\mathbf{x}+\mathbf{H}\mathbf{p})d\mathbf{p} \\
&= o_p(1)
\end{aligned}
$$

■

Lemma 2. *Let*

$$
W_{2n}(\mathbf{x},\mathbf{t}) = \frac{1}{n}\sum_{i=1}^{n} K_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{x}\right)K_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{t}\right).
$$

For any $\mathbf{x}\in D$, under assumptions (A1), (A3), (A7) and (A9), then

$$
W_{2n}(\mathbf{x},\mathbf{t}) = |\mathbf{H}|^{-1}K^{(2)}(\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t}))f(\mathbf{t})\cdot\{1+o_p(1)\}.
$$

Proof of Lemma 2. For any $\mathbf{x},\mathbf{t}\in D$

$$
\begin{aligned}
\mathbb{E}(|\mathbf{H}|W_{2n}(\mathbf{x},\mathbf{t})) &= |\mathbf{H}|\int K_{\mathbf{H}}\left(\mathbf{u}-\mathbf{x}\right)K_{\mathbf{H}}\left(\mathbf{u}-\mathbf{t}\right)f(\mathbf{u})d\mathbf{u} \\
&= \int K\left(\mathbf{p}\right)K\left(\mathbf{p}-\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})\right)f(\mathbf{t}+\mathbf{H}\mathbf{p})d\mathbf{p} \\
&= K^{(2)}(\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t}))\{f(\mathbf{t})+o(1)\}.
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
\mathrm{Var}(|\mathbf{H}|W_{2n}(\mathbf{x},\mathbf{t})) &\leq \frac{|\mathbf{H}|^2}{n}\int K_{\mathbf{H}}^{2}\left(\mathbf{u}-\mathbf{x}\right)K_{\mathbf{H}}^{2}\left(\mathbf{u}-\mathbf{t}\right)f(\mathbf{u})d\mathbf{u} \\
&= \frac{1}{n|\mathbf{H}|}\int K^{2}\left(\mathbf{p}\right)K^{2}\left(\mathbf{p}-\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})\right)f(\mathbf{t}+\mathbf{H}\mathbf{p})d\mathbf{p} \\
&= o_p(1).
\end{aligned}
$$

■

Lemma 3. *Let*

$$
W_{3n}(\mathbf{x},\mathbf{t}) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K_{\mathbf{H}}^{2}\left(\mathbf{X}_i-\mathbf{x}\right)K_{\mathbf{H}}^{2}\left(\mathbf{X}_j-\mathbf{t}\right)\rho_n^2(\mathbf{X}_i-\mathbf{X}_j).
$$

For any $\mathbf{x},\mathbf{t}\in D$, under assumptions (A1), (A3) and (A9), then

$$
\mathbb{E}(W_{3n}(\mathbf{x},\mathbf{t})) = |\mathbf{H}|^{-2}f(\mathbf{x})f(\mathbf{t})\int\int K^2(\mathbf{p})K^2(\mathbf{q})\rho_n^2(\mathbf{x}-\mathbf{t}+\mathbf{H}(\mathbf{p}-\mathbf{q}))d\mathbf{p}d\mathbf{q}\cdot\{1+o(1)\}.
$$

Proof of Lemma 3. For any $\mathbf{x},\mathbf{t}\in D$

$$
\begin{aligned}
\mathbb{E}(W_{3n}(\mathbf{x},\mathbf{t})) &= \int\int K_{\mathbf{H}}^{2}\left(\mathbf{u}-\mathbf{x}\right)K_{\mathbf{H}}^{2}\left(\mathbf{v}-\mathbf{t}\right)\rho_n^2(\mathbf{u}-\mathbf{v})f(\mathbf{u})f(\mathbf{v})d\mathbf{u}d\mathbf{v} \\
&= |\mathbf{H}|^{-2}\int\int K^2(\mathbf{p})K^2(\mathbf{q})\rho_n^2(\mathbf{x}-\mathbf{t}+\mathbf{H}(\mathbf{p}-\mathbf{q}))f(\mathbf{x}+\mathbf{H}\mathbf{p})f(\mathbf{t}+\mathbf{H}\mathbf{q})d\mathbf{p}d\mathbf{q} \\
&= |\mathbf{H}|^{-2}f(\mathbf{x})f(\mathbf{t})\int\int K^2(\mathbf{p})K^2(\mathbf{q})\rho_n^2(\mathbf{x}-\mathbf{t}+\mathbf{H}(\mathbf{p}-\mathbf{q}))d\mathbf{p}d\mathbf{q}\cdot\{1+o(1)\}.
\end{aligned}
$$

■

Lemma 4. *Let*

$$W_{4n}(\mathbf{x},\mathbf{t}) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{t}) \rho_n(\mathbf{X}_i - \mathbf{X}_j).$$

For any $\mathbf{x},\mathbf{t} \in D$*, under assumptions (A1), (A3) and (A9), then*

$$\mathbb{E}(W_{4n}(\mathbf{x},\mathbf{t})) = f(\mathbf{x})f(\mathbf{t})\int\int K(\mathbf{p}) K(\mathbf{q}) \rho_n(\mathbf{x}-\mathbf{t}+\mathbf{H}(\mathbf{p}-\mathbf{q}))d\mathbf{p}d\mathbf{q} + o(1).$$

Proof of Lemma 4. For any $\mathbf{x},\mathbf{t} \in D$,

$$\begin{aligned}
\mathbb{E}(W_{4n}(\mathbf{x},\mathbf{t})) &= \int\int K_{\mathbf{H}}(\mathbf{u}-\mathbf{x}) K_{\mathbf{H}}(\mathbf{v}-\mathbf{t}) \rho_n(\mathbf{u}-\mathbf{v}) f(\mathbf{u})f(\mathbf{v})d\mathbf{u}d\mathbf{v} \\
&= \int\int K(\mathbf{p}) K(\mathbf{q}) \rho_n(\mathbf{x}-\mathbf{t}+\mathbf{H}(\mathbf{p}-\mathbf{q}))f(\mathbf{x}+\mathbf{H}\mathbf{p})f(\mathbf{t}+\mathbf{H}\mathbf{q})d\mathbf{p}d\mathbf{q} \\
&= f(\mathbf{x})f(\mathbf{t})\int\int K(\mathbf{p}) K(\mathbf{q}) \rho_n(\mathbf{x}-\mathbf{t}+\mathbf{H}(\mathbf{p}-\mathbf{q}))d\mathbf{p}d\mathbf{q} + o(1).
\end{aligned}$$

■

Lemma 5. *Let*

$$W_{5n}(\mathbf{x},\mathbf{t}) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{t}) K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{t}) \rho_n^2(\mathbf{X}_i - \mathbf{X}_j).$$

For any $\mathbf{x},\mathbf{t} \in D$*, under assumptions (A1), (A3) and (A9), then*

$$\begin{aligned}
\mathbb{E}(W_{5n}(\mathbf{x},\mathbf{t})) &= |\mathbf{H}|^{-2} f^2(\mathbf{t})\int\int K\left(-\mathbf{p}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})\right) K\left(-\mathbf{q}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})\right) \\
&\cdot\ K(\mathbf{p})K(\mathbf{q})\rho_n^2(\mathbf{H}(\mathbf{p}-\mathbf{q}))d\mathbf{p}d\mathbf{q}\cdot\{1+o(1)\}.
\end{aligned}$$

Proof of Lemma 5. For any $\mathbf{x},\mathbf{t} \in D$,

$$\begin{aligned}
\mathbb{E}(W_{5n}(\mathbf{x},\mathbf{t})) &= \int\int K_{\mathbf{H}}(\mathbf{u}-\mathbf{x}) K_{\mathbf{H}}(\mathbf{u}-\mathbf{t}) K_{\mathbf{H}}(\mathbf{v}-\mathbf{x}) K_{\mathbf{H}}(\mathbf{v}-\mathbf{t}) \rho_n^2(\mathbf{u}-\mathbf{v}) \\
&\cdot\ f(\mathbf{u})f(\mathbf{v})d\mathbf{u}d\mathbf{v} \\
&= |\mathbf{H}|^{-2}\int\int K\left(-\mathbf{p}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})\right) K\left(-\mathbf{q}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})\right) K(\mathbf{p}) K(\mathbf{q}) \\
&\cdot\ f(\mathbf{t}+\mathbf{H}\mathbf{p})f(\mathbf{t}+\mathbf{H}\mathbf{q})\rho_n^2(\mathbf{H}(\mathbf{p}-\mathbf{q}))d\mathbf{p}d\mathbf{q} \\
&= |\mathbf{H}|^{-2} f^2(\mathbf{t})\int\int K\left(-\mathbf{p}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})\right) K\left(-\mathbf{q}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})\right) \\
&\cdot\ K(\mathbf{p}) K(\mathbf{q}) \rho_n^2(\mathbf{H}(\mathbf{p}-\mathbf{q}))d\mathbf{p}d\mathbf{q}\cdot\{1+o(1)\}.
\end{aligned}$$

■

Lemma 6. *Let*

$$W_{6n}(\mathbf{x},\mathbf{t}) = \frac{1}{n^3}\sum_{i\neq j}\sum_{k\neq i,j} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{t}) K_{\mathbf{H}}(\mathbf{X}_k - \mathbf{t})\rho_n(\mathbf{X}_i - \mathbf{X}_k)\rho_n(\mathbf{X}_j - \mathbf{X}_i).$$

For any $\mathbf{x},\mathbf{t} \in D$*, under assumptions (A1), (A3) and (A9), then*

$$\begin{aligned}
\mathbb{E}(W_{6n}(\mathbf{x},\mathbf{t})) &= |\mathbf{H}|^{-1} f^2(\mathbf{x})f(\mathbf{t})\int\int\int K(\mathbf{p})K(\mathbf{q}) K\left(-\mathbf{p}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})\right) K(\mathbf{r}) \\
&\cdot\ \rho_n(\mathbf{H}(\mathbf{p}-\mathbf{q}))\rho_n(\mathbf{x}-\mathbf{t}+\mathbf{H}(\mathbf{p}-\mathbf{r}))d\mathbf{p}d\mathbf{q}d\mathbf{r}\cdot\{1+o(1)\}.
\end{aligned}$$

Proof of Lemma 6. For any $\mathbf{x},\mathbf{t}\in D$,

$$\begin{aligned}
\mathbb{E}(W_{6n}(\mathbf{x},\mathbf{t})) &= \int\int\int K_{\mathbf{H}}\left(\mathbf{u}-\mathbf{x}\right)K_{\mathbf{H}}\left(\mathbf{v}-\mathbf{x}\right)K_{\mathbf{H}}\left(\mathbf{u}-\mathbf{t}\right)K_{\mathbf{H}}\left(\mathbf{y}-\mathbf{t}\right)\\
&\cdot\ \rho_n(\mathbf{u}-\mathbf{y})\rho_n(\mathbf{u}-\mathbf{v})f(\mathbf{u})f(\mathbf{v})f(\mathbf{y})d\mathbf{u}d\mathbf{v}d\mathbf{y}\\
&= |\mathbf{H}|^{-1}\int\int\int K(\mathbf{p})K\left(\mathbf{q}\right)K\left(-\mathbf{p}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})\right)K\left(\mathbf{r}\right)\\
&\cdot\ f(\mathbf{x}+\mathbf{H}\mathbf{p})f(\mathbf{x}+\mathbf{H}\mathbf{q})f(\mathbf{t}+\mathbf{H}\mathbf{r})\rho_n(\mathbf{H}(\mathbf{p}-\mathbf{q}))\rho_n(\mathbf{x}-\mathbf{t}+\mathbf{H}(\mathbf{p}-\mathbf{r}))d\mathbf{p}d\mathbf{q}d\mathbf{r}\\
&= |\mathbf{H}|^{-1}f^2(\mathbf{x})f(\mathbf{t})\int\int\int K(\mathbf{p})K\left(\mathbf{q}\right)K\left(-\mathbf{p}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})\right)K\left(\mathbf{r}\right)\\
&\cdot\ \rho_n(\mathbf{H}(\mathbf{p}-\mathbf{q}))\rho_n(\mathbf{x}-\mathbf{t}+\mathbf{H}(\mathbf{p}-\mathbf{r}))d\mathbf{p}d\mathbf{q}d\mathbf{r}\cdot\{1+o(1)\}.
\end{aligned}$$

∎

Next, the proof of Theorem 1 of the main paper is presented.

Theorem 1. *Under Assumptions* (A1)-(A10)*, and if* $0<V<\infty$*, it can be proved that*

$$V^{-1/2}(T_n-b_{0\mathbf{H}}-b_{1\mathbf{H}})\to_{\mathcal{L}} N(0,1)\ \ as\ n\to\infty,$$

where $\to_{\mathcal{L}}$ *denotes convergence in distribution, with*

$$\begin{aligned}
b_{0\mathbf{H}} &= |\mathbf{H}|^{-1/2}\sigma^2K^{(2)}(\mathbf{0})\Big[\int\frac{w(\mathbf{x})}{f(\mathbf{x})}d\mathbf{x}+\rho_c\int w(\mathbf{x})d\mathbf{x}\Big],\\
b_{1\mathbf{H}} &= \int(K_{\mathbf{H}}*g(\mathbf{x}))^2w(\mathbf{x})d\mathbf{x},
\end{aligned}$$

and

$$V = 2\sigma^4K^{(4)}(\mathbf{0})\Big[\int\frac{w^2(\mathbf{x})}{f^2(\mathbf{x})}d\mathbf{x}+2\rho_c\int\frac{w^2(\mathbf{x})}{f(\mathbf{x})}d\mathbf{x}+4\rho_c^2\int w^2(\mathbf{x})d\mathbf{x}\Big],$$

where $K^{(j)}$ *denotes the* j*-times convolution product of* K *with itself.*

Proof of Theorem 1. The test statistic (1) can be written as

$$\begin{aligned}
T_n &= n|\mathbf{H}|^{1/2}\int(\hat{m}^{LL}_{\mathbf{H}}(\mathbf{x})-\hat{m}^{LL}_{\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x}))^2w(\mathbf{x})d\mathbf{x}\\
&= n|\mathbf{H}|^{1/2}\int\left[e_1'\left(\frac{1}{n}X_x'W_xX_x\right)^{-1}\frac{1}{n}\sum_{i=1}^n(1,(\mathbf{X}_i-\mathbf{x})')K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})(Z_i-m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i))\right]^2w(\mathbf{x})d\mathbf{x}\\
&= n|\mathbf{H}|^{1/2}\int\left[e_1'\begin{pmatrix}\frac{1}{n}\sum_{i=1}^nK_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x}) & \frac{1}{n}\sum_{i=1}^nK_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})'\\ \frac{1}{n}\sum_{i=1}^nK_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x}) & \frac{1}{n}\sum_{i=1}^nK_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})'\end{pmatrix}^{-1}\right.\\
&\cdot\left.\begin{pmatrix}\frac{1}{n}\sum_{i=1}^nK_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})(Z_i-m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i))\\ \frac{1}{n}\sum_{i=1}^nK_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})(Z_i-m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i))\end{pmatrix}\right]^2w(\mathbf{x})d\mathbf{x}.
\end{aligned}$$

According to Liu (2001) and taking into account that for every $\eta>0$, $\hat{f}_{\mathbf{H}}(\mathbf{x})=\frac{1}{n}\sum_{i=1}^nK_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})=f(\mathbf{x})+O_p(n^{-2/(4+d)+\eta})$ uniformly in $\mathbf{x}$ (see Härdle and Mammen 1993), it follows that

$$\begin{aligned}
\frac{1}{n}X_x'W_xX_x &= \begin{pmatrix}\frac{1}{n}\sum_{i=1}^nK_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x}) & \frac{1}{n}\sum_{i=1}^nK_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})'\\ \frac{1}{n}\sum_{i=1}^nK_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x}) & \frac{1}{n}\sum_{i=1}^nK_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})'\end{pmatrix}\\
&= \begin{pmatrix}f(\mathbf{x})+O_p(n^{-2/(4+d)+\eta}) & \mu_2(K)\nabla f(\mathbf{x})'\mathbf{H}^2+O_p(n^{-2/(4+d)+\eta}\mathbf{H}^2)\\ \mu_2(K)\mathbf{H}^2\nabla f(\mathbf{x})+O_p(n^{-2/(4+d)+\eta}\mathbf{H}^2) & \mu_2(K)f(\mathbf{x})\mathbf{H}^2+O_p(\mathbf{H}\mathbf{1}_{d\times d}\mathbf{H})\end{pmatrix},
\end{aligned}$$

where $\nabla f(\mathbf{x})$ denotes the $d \times 1$ vector of first-order partial derivatives of f (and $\nabla f(\mathbf{x})'$ its transpose).

Therefore,

$$
\begin{aligned}
T_n &= n|\mathbf{H}|^{1/2} \\
&\cdot \int \Bigg[e_1' \begin{pmatrix} f(\mathbf{x}) + O_p(n^{-2/(4+d)+\eta}) & \mu_2(K)\nabla f(\mathbf{x})'\mathbf{H}^2 + O_p(n^{-2/(4+d)+\eta}\mathbf{H}^2) \\ \mu_2(K)\mathbf{H}^2\nabla f(\mathbf{x}) + O_p(n^{-2/(4+d)+\eta}\mathbf{H}^2) & \mu_2(K) f(\mathbf{x})\mathbf{H}^2 + O_p(\mathbf{H}\mathbf{1}_{d\times d}\mathbf{H}) \end{pmatrix}^{-1} \\
&\cdot \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(Z_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)) \\ \frac{1}{n}\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})(Z_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)) \end{pmatrix} \Bigg]^2 w(\mathbf{x})d\mathbf{x} \\
&= n|\mathbf{H}|^{1/2} \\
&\cdot \int \Bigg[e_1' \begin{pmatrix} f^{-1}(\mathbf{x}) + O_p(n^{-2/(4+d)+\eta}) & -f^{-2}(\mathbf{x})\nabla f(\mathbf{x})' + O_p(n^{-2/(4+d)+\eta}\mathbf{1}_d') \\ -f^{-2}(\mathbf{x})\nabla f(\mathbf{x}) + O_p(n^{-2/(4+d)+\eta}\mathbf{1}) & \{\mu_2(K) f(\mathbf{x})\mathbf{H}^2\}^{-1} + O_p(n^{-2/(4+d)+\eta}\mathbf{H}\mathbf{1}_{d\times d}\mathbf{H}) \end{pmatrix} \\
&\cdot \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(Z_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)) \\ \frac{1}{n}\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})(Z_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)) \end{pmatrix} \Bigg]^2 w(\mathbf{x})d\mathbf{x} \\
&= n|\mathbf{H}|^{1/2} \int \Bigg[\frac{1}{nf(\mathbf{x})} \sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(Z_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)) \\
&- \nabla f(\mathbf{x}) \frac{1}{nf^2(\mathbf{x})} \sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})(Z_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)) \Bigg]^2 w(\mathbf{x})d\mathbf{x} + O_p(n^{-2/(4+d)+\eta}) \\
&= T_{n1} + T_{n2} + 2T_{n12} + O_p(n^{-2/(4+d)+\eta}), \qquad (2)
\end{aligned}
$$

with

$$
\begin{aligned}
T_{n1} &= n|\mathbf{H}|^{1/2} \int \left[\frac{1}{nf(\mathbf{x})} \sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(Z_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)) \right]^2 w(\mathbf{x})d\mathbf{x}, \\
T_{n2} &= n|\mathbf{H}|^{1/2} \int \left[\nabla f(\mathbf{x}) \frac{1}{nf^2(\mathbf{x})} \sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})(Z_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)) \right]^2 w(\mathbf{x})d\mathbf{x},
\end{aligned}
$$

and the T_{n12} term is the integral of the cross product.

Regarding T_{n1}, taking into account that the regression functions considered are of the form $m = m_{\boldsymbol{\beta}_0} + n^{-1/2}|\mathbf{H}|^{-1/4}g$ (see the main paper), one gets

$$
\begin{aligned}
T_{n1} &= n|\mathbf{H}|^{1/2} \int \left[\frac{1}{nf(\mathbf{x})} \sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(Z_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)) \right]^2 w(\mathbf{x})d\mathbf{x} \\
&= n|\mathbf{H}|^{1/2} \int \left[\frac{1}{nf(\mathbf{x})} \sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(m(\mathbf{X}_i) + \varepsilon_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)) \right]^2 w(\mathbf{x})d\mathbf{x} \\
&= n|\mathbf{H}|^{1/2} \int \left[\frac{1}{nf(\mathbf{x})} \sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(m_{\boldsymbol{\beta}_0}(\mathbf{X}_i) + n^{-1/2}|\mathbf{H}|^{-1/4}g(\mathbf{X}_i) + \varepsilon_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)) \right]^2 w(\mathbf{x})d\mathbf{x} \\
&= n|\mathbf{H}|^{1/2} \int \frac{1}{f^2(\mathbf{x})} (I_1(\mathbf{x}) + I_2(\mathbf{x}) + I_3(\mathbf{x}))^2 w(\mathbf{x})d\mathbf{x},
\end{aligned}
$$

where

$$\begin{aligned}
I_1(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})\,(m_{\boldsymbol{\beta}_0}(\mathbf{X}_i)-m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)),\\
I_2(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})\,n^{-1/2}|\mathbf{H}|^{-1/4}g(\mathbf{X}_i),\\
I_3(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})\,\varepsilon_i.
\end{aligned}$$

With respect to the term $I_1(\mathbf{x})$, using assumptions (A1)–(A3) and (A7), and given that the difference $m_{\hat{\boldsymbol{\beta}}}(\mathbf{x})-m_{\boldsymbol{\beta}_0}(\mathbf{x})=O_p(n^{-1/2})$ uniformly in $\mathbf{x}$ (see the main paper), it is obtained that

$$\begin{aligned}
& n|\mathbf{H}|^{1/2}\int\frac{1}{f^2(\mathbf{x})}I_1^2(\mathbf{x})w(\mathbf{x})d\mathbf{x}\\
=\ & n|\mathbf{H}|^{1/2}\int\frac{1}{f^2(\mathbf{x})}\left[\frac{1}{n}\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})\,(m_{\boldsymbol{\beta}_0}(\mathbf{X}_i)-m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i))\right]^2 w(\mathbf{x})d\mathbf{x}\\
=\ & O_p(|\mathbf{H}|^{1/2}). \qquad (3)
\end{aligned}$$

As for the term $I_2(\mathbf{x})$, taking into account Lemma 1, it follows that

$$\begin{aligned}
& n|\mathbf{H}|^{1/2}\int\frac{1}{f^2(\mathbf{x})}I_2^2(\mathbf{x})w(\mathbf{x})d\mathbf{x}\\
=\ & n|\mathbf{H}|^{1/2}\int\frac{1}{f^2(\mathbf{x})}\left[\frac{1}{n}\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})\,n^{-1/2}|\mathbf{H}|^{-1/4}g(\mathbf{X}_i)\right]^2 w(\mathbf{x})d\mathbf{x}\\
=\ & \int\frac{1}{f^2(\mathbf{x})}\left[\int K(\mathbf{p})\,g(\mathbf{x}+\mathbf{H}\mathbf{p})f(\mathbf{x}+\mathbf{H}\mathbf{p})d\mathbf{p}+o_p(1)\right]^2 w(\mathbf{x})d\mathbf{x}\\
=\ & \int\frac{1}{f^2(\mathbf{x})}\left[\int K(\mathbf{p})\,g(\mathbf{x}+\mathbf{H}\mathbf{p})\{f(\mathbf{x})+o(1)\}d\mathbf{p}\right]^2 w(\mathbf{x})d\mathbf{x}\cdot\{1+o_p(1)\}\\
=\ & \int\left[\int K_{\mathbf{H}}(\mathbf{u}-\mathbf{x})\,g(\mathbf{u})d\mathbf{u}\right]^2 w(\mathbf{x})d\mathbf{x}\cdot\{1+o_p(1)\}\\
=\ & \int(K_{\mathbf{H}}*g)^2(\mathbf{x})w(\mathbf{x})d\mathbf{x}\cdot\{1+o_p(1)\}. \qquad (4)
\end{aligned}$$

The leading term of (4) is the term $b_{1\mathbf{H}}$ in Theorem 1. Finally, the term $I_3(\mathbf{x})$, associated with the error component of the model, can be split as

$$\begin{aligned}
n|\mathbf{H}|^{1/2}\int\frac{1}{f^2(\mathbf{x})}I_3^2(\mathbf{x})w(\mathbf{x})d\mathbf{x} &= n|\mathbf{H}|^{1/2}\int\frac{1}{f^2(\mathbf{x})}\left[\frac{1}{n}\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})\,\varepsilon_i\right]^2 w(\mathbf{x})d\mathbf{x}\\
&= n|\mathbf{H}|^{1/2}\int\frac{1}{f^2(\mathbf{x})}\frac{1}{n^2}\sum_{i=1}^{n} K_{\mathbf{H}}^2(\mathbf{X}_i-\mathbf{x})\,\varepsilon_i^2 w(\mathbf{x})d\mathbf{x}\\
&+ n|\mathbf{H}|^{1/2}\int\frac{1}{f^2(\mathbf{x})}\frac{1}{n^2}\sum_{i\neq j} K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})\,K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})\,\varepsilon_i\varepsilon_j w(\mathbf{x})d\mathbf{x}\\
&= I_{31}+I_{32}.
\end{aligned}$$

Close expressions of I_{31} and I_{32} can be obtained computing the expectation and the variance of these terms. For doing so, general results on the conditional expectation and conditional variance can be used. Specifically, given two random variables X and Y, it is known that $\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|Y))$ and $\text{Var}(X) = \mathbb{E}(\text{Var}(X|Y)) + \text{Var}(\mathbb{E}(X|Y))$.

For I_{31}, using the result for the conditional mean, it follows that $\mathbb{E}(I_{31}) = \mathbb{E}(\mathbb{E}(I_{31}|\mathbf{X}_1, \ldots, \mathbf{X}_n))$. Firstly,

$$\begin{aligned}\mathbb{E}(|\mathbf{H}|^{1/2} I_{31}|\mathbf{X}_1, \ldots, \mathbf{X}_n) &= \mathbb{E}\left[n|\mathbf{H}| \int \frac{1}{f^2(\mathbf{x})} \frac{1}{n^2} \sum_{i=1}^{n} K_{\mathbf{H}}^2 (\mathbf{X}_i - \mathbf{x})\, \varepsilon_i^2 w(\mathbf{x}) d\mathbf{x} |\mathbf{X}_1, \ldots, \mathbf{X}_n\right] \\ &= \sigma^2 n|\mathbf{H}| \int \frac{1}{f^2(\mathbf{x})} \frac{1}{n^2} \sum_{i=1}^{n} K_{\mathbf{H}}^2 (\mathbf{X}_i - \mathbf{x})\, w(\mathbf{x}) d\mathbf{x}. \end{aligned} \tag{5}$$

Considering the first part of the proof of Lemma 2, one gets that,

$$\begin{aligned}\mathbb{E}(|\mathbf{H}|^{1/2} I_{31}) = \mathbb{E}(\mathbb{E}(|\mathbf{H}|^{1/2} I_{31}|\mathbf{X}_1, \ldots, \mathbf{X}_n)) &= \mathbb{E}\left[\sigma^2 n|\mathbf{H}| \int \frac{1}{f^2(\mathbf{x})} \frac{1}{n^2} \sum_{i=1}^{n} K_{\mathbf{H}}^2 (\mathbf{X}_i - \mathbf{x})\, w(\mathbf{x}) d\mathbf{x}\right] \\ &= \sigma^2 |\mathbf{H}| \int \frac{1}{f^2(\mathbf{x})} |\mathbf{H}|^{-1} K^{(2)}(\mathbf{0}) \{f(\mathbf{x}) + o(1)\} w(\mathbf{x}) d\mathbf{x} \\ &= \sigma^2 K^{(2)}(\mathbf{0}) \int \frac{w(\mathbf{x})}{f(\mathbf{x})} d\mathbf{x} \cdot \{1 + o(1)\}. \end{aligned} \tag{6}$$

On the other hand,

$$\text{Var}(I_{31}) = \mathbb{E}(\text{Var}(I_{31}|\mathbf{X}_1, \ldots, \mathbf{X}_n)) + \text{Var}(\mathbb{E}(I_{31}|\mathbf{X}_1, \ldots, \mathbf{X}_n)). \tag{7}$$

Using assumption (A5), it is obtained that

$$\begin{aligned} & \text{Var}(|\mathbf{H}|^{1/2} I_{31}|\mathbf{X}_1, \ldots, \mathbf{X}_n) \\ = \; & \text{Var}\left[n|\mathbf{H}| \int \frac{1}{f^2(\mathbf{x})} \frac{1}{n^2} \sum_{i=1}^{n} K_{\mathbf{H}}^2 (\mathbf{X}_i - \mathbf{x})\, \varepsilon_i^2 w(\mathbf{x}) d\mathbf{x} |\mathbf{X}_1, \ldots, \mathbf{X}_n\right] \\ = \; & \frac{1}{n^2} |\mathbf{H}|^2 \sum_{i=1}^{n} \sum_{j=1}^{n} \int \int \frac{1}{f^2(\mathbf{x}) f^2(\mathbf{t})} K_{\mathbf{H}}^2 (\mathbf{X}_i - \mathbf{x})\, K_{\mathbf{H}}^2 (\mathbf{X}_j - \mathbf{t})\, w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \\ \cdot \; & \text{Cov}(\varepsilon_i^2, \varepsilon_j^2) \\ = \; & \frac{1}{n^2} |\mathbf{H}|^2 \sum_{i=1}^{n} \sum_{j=1}^{n} \int \int \frac{1}{f^2(\mathbf{x}) f^2(\mathbf{t})} K_{\mathbf{H}}^2 (\mathbf{X}_i - \mathbf{x})\, K_{\mathbf{H}}^2 (\mathbf{X}_j - \mathbf{t})\, w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \\ \cdot \; & 2(\text{Cov}(\varepsilon_i, \varepsilon_j))^2 \\ = \; & \frac{2\sigma^4}{n^2} |\mathbf{H}|^2 \sum_{i=1}^{n} \sum_{j=1}^{n} \int \int \frac{1}{f^2(\mathbf{x}) f^2(\mathbf{t})} K_{\mathbf{H}}^2 (\mathbf{X}_i - \mathbf{x})\, K_{\mathbf{H}}^2 (\mathbf{X}_j - \mathbf{t})\, w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \\ \cdot \; & \rho_n^2 (\mathbf{X}_i - \mathbf{X}_j) \end{aligned}$$

and, therefore by using and Lemma 3,

$$
\begin{aligned}
& \mathbb{E}(\mathrm{Var}(|\mathbf{H}|^{1/2} I_{31}|\mathbf{X}_1,\ldots,\mathbf{X}_n)) \\
= & \mathbb{E}\Bigg[\frac{2\sigma^4}{n^2}|\mathbf{H}|^2\sum_{i=1}^n\sum_{j=1}^n\int\int\frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})}K^2_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{x}\right)K^2_{\mathbf{H}}\left(\mathbf{X}_j-\mathbf{t}\right)w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t} \\
\cdot & \rho_n^2(\mathbf{X}_i-\mathbf{X}_j)\Bigg] \\
= & 2\sigma^4|\mathbf{H}|^2\int\int\frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})}|\mathbf{H}|^{-2}\int\int K^2(\mathbf{p})K^2(\mathbf{q})\rho_n^2(\mathbf{x}-\mathbf{t}+\mathbf{H}(\mathbf{p}-\mathbf{q}))d\mathbf{p}d\mathbf{q} \\
\cdot & f(\mathbf{x})f(\mathbf{t})w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t}\cdot\{1+o(1)\} \\
= & 2\sigma^4\int\int\int\int\frac{K^2(\mathbf{p})K^2(\mathbf{q})}{f(\mathbf{x})f(\mathbf{t})}w(\mathbf{x})w(\mathbf{t})\rho_n^2(\mathbf{x}-\mathbf{t}+\mathbf{H}(\mathbf{p}-\mathbf{q}))d\mathbf{p}d\mathbf{q}d\mathbf{x}d\mathbf{t} \\
\cdot & \{1+o(1)\} \\
= & 2\sigma^4|\mathbf{H}|\int\int\int\int\frac{K^2(\mathbf{p})K^2(\mathbf{q})}{f(\mathbf{x})f(\mathbf{x}+\mathbf{H}\mathbf{u})}w(\mathbf{x})w(\mathbf{x}+\mathbf{H}\mathbf{u})\rho_n^2(\mathbf{H}(\mathbf{p}-\mathbf{q}+\mathbf{u}))d\mathbf{p}d\mathbf{q}d\mathbf{x}d\mathbf{u} \\
\cdot & \{1+o(1)\} \\
= & 2\sigma^4|\mathbf{H}|\int\int\int\int\frac{K^2(\mathbf{p})K^2(\mathbf{q})}{f^2(\mathbf{x})}w^2(\mathbf{x})\rho_n^2(\mathbf{H}(\mathbf{p}-\mathbf{q}+\mathbf{u}))d\mathbf{p}d\mathbf{q}d\mathbf{x}d\mathbf{u} \\
\cdot & \{1+o(1)\}.
\end{aligned}
$$

Let

$$
j_n(\mathbf{p},\mathbf{u}) \;=\; n|\mathbf{H}|\int K^2(\mathbf{q})\rho_n^2(\mathbf{H}(\mathbf{p}-\mathbf{q}+\mathbf{u}))d\mathbf{q}.
$$

Notice that, using assumption (A4),

$$
\begin{aligned}
|j_n(\mathbf{p},\mathbf{u})| &\leq K_M^2(n|\mathbf{H}|\int|\rho_n^2(\mathbf{H}(\mathbf{p}-\mathbf{q}+\mathbf{u}))|d\mathbf{q}) \\
&\leq K_M^2(n\int|\rho_n(\mathbf{t})|d\mathbf{t}) \\
&\leq K_M^2\rho_M,
\end{aligned}
$$

where $K_M\equiv\max_{\mathbf{x}}(K(\mathbf{x}))$ and $\rho_M\equiv\max_{\mathbf{x}}(\rho(\mathbf{x}))$, and using assumptions (A2), (A3), (A7) and (A9), one gets that

$$
\mathbb{E}(\mathrm{Var}(I_{31}|\mathbf{X}_1,\ldots,\mathbf{X}_n)) = o_p(1). \tag{8}
$$

On the other hand, using expression (5), the second part of Lemma 2 and assumption (A9), it follows that

$$
\begin{aligned}
& \mathrm{Var}(\mathbb{E}(|\mathbf{H}|^{1/2}I_{31}|\mathbf{X}_1,\ldots,\mathbf{X}_n)) \\
= & \mathrm{Var}\left[\sigma^2 n|\mathbf{H}|\int\frac{1}{f^2(\mathbf{x})}\frac{1}{n^2}\sum_{i=1}^n K^2_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{x}\right)w(\mathbf{x})d\mathbf{x}\right] \\
= & \sum_{i=1}^n\mathrm{Var}\left[\sigma^2 n|\mathbf{H}|\int\frac{1}{f^2(\mathbf{x})}\frac{1}{n^2}K^2_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{x}\right)w(\mathbf{x})d\mathbf{x}\right] \\
\leq & \sigma^4|\mathbf{H}|^2\sum_{i=1}^n\mathbb{E}\left[\int\int\frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})}\frac{1}{n^2}K^2_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{x}\right)K^2_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{t}\right)w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t}\right] \\
= & o_p(1).
\end{aligned} \tag{9}
$$

Now, considering (7), (8) and (9), it is obtained that

$$\mathrm{Var}(|\mathbf{H}|^{1/2} I_{31}) = o_p(1), \tag{10}$$

and considering (6) and (10),

$$I_{31} = \sigma^2 |\mathbf{H}|^{-1/2} K^{(2)}(\mathbf{0}) \int \frac{w(\mathbf{x})}{f(\mathbf{x})} d\mathbf{x} \cdot \{1 + o_p(1))\}. \tag{11}$$

Taking into account assumption (A9), the leading term of (11) corresponds to the first term of $b_{0\mathbf{H}}$ in Theorem 1 of the main paper.

Now, consider the term

$$I_{32} \;=\; n|\mathbf{H}|^{1/2} \int \frac{1}{f^2(\mathbf{x})} \frac{1}{n^2} \sum_{i \neq j} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x}) \varepsilon_i \varepsilon_j w(\mathbf{x}) d\mathbf{x}.$$

Let

$$\kappa_{ij} \;=\; n|\mathbf{H}|^{1/2} \int \frac{1}{f^2(\mathbf{x})} \frac{1}{n^2} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x}) \varepsilon_i \varepsilon_j w(\mathbf{x}) d\mathbf{x},$$

thus,

$$I_{32} = \sum_{i \neq j} \kappa_{ij},$$

and this can be seen as a U-statistic with degenerate kernel. To obtain the asymptotic normality of I_{32}, Theorem 2 given in Kim et al. (2013) will be applied. In this work, the central limit theorem for degenerate reduced U-statistics under $\alpha-$mixing is derived. The assumptions of this result hold (specifically, assumption (A6)) and the expectation and the variance of I_{32} should be computed.

Proceeding as for I_{31}, it follows that $\mathbb{E}(I_{32}) = \mathbb{E}(\mathbb{E}(I_{32}|\mathbf{X}_1, \ldots, \mathbf{X}_n))$. Taking into account the first part of Lemma 4, one gets that

$$\begin{aligned}
\mathbb{E}(I_{32}|\mathbf{X}_1, \ldots, \mathbf{X}_n) &= \mathbb{E}\left[n|\mathbf{H}|^{1/2} \frac{1}{n^2} \sum_{i \neq j} \int \frac{1}{f^2(\mathbf{x})} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x}) w(\mathbf{x}) d\mathbf{x} \cdot \varepsilon_i \varepsilon_j | \mathbf{X}_1, \ldots, \mathbf{X}_n \right] \\
&= n|\mathbf{H}|^{1/2} \frac{1}{n^2} \int \frac{1}{f^2(\mathbf{x})} \sum_{i \neq j} \mathbb{E}(\varepsilon_i \varepsilon_j) K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x}) w(\mathbf{x}) d\mathbf{x} \\
&= n|\mathbf{H}|^{1/2} \int \frac{1}{f^2(\mathbf{x})} \frac{1}{n^2} \sum_{i \neq j} \mathrm{Cov}(\varepsilon_i, \varepsilon_j) K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x}) w(\mathbf{x}) d\mathbf{x} \\
&= |\mathbf{H}|^{1/2} \sigma^2 \int \frac{1}{f^2(\mathbf{x})} \frac{1}{n} \sum_{i \neq j} \rho_n(\mathbf{X}_i - \mathbf{X}_j) K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x}) w(\mathbf{x}) d\mathbf{x}, \qquad (12)
\end{aligned}$$

and, therefore,

$$\begin{aligned}
& \mathbb{E}(\mathbb{E}(I_{32}|\mathbf{X}_1, \ldots, \mathbf{X}_n)) \\
= \;& \mathbb{E}\left[n|\mathbf{H}|^{1/2} \sigma^2 \int \frac{1}{f^2(\mathbf{x})} \frac{1}{n^2} \sum_{i \neq j} \rho_n(\mathbf{X}_i - \mathbf{X}_j) K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x}) w(\mathbf{x}) d\mathbf{x} \right] \\
= \;& n|\mathbf{H}|^{1/2} \sigma^2 \int \frac{1}{f^2(\mathbf{x})} \left(\frac{n-1}{n} f^2(\mathbf{x}) \int \int K(\mathbf{p}) K(\mathbf{q}) \rho_n(\mathbf{H}(\mathbf{p} - \mathbf{q})) d\mathbf{p} d\mathbf{q} + o(1) \right) \\
\cdot \;& w(\mathbf{x}) d\mathbf{x} \\
= \;& \frac{n-1}{n} |\mathbf{H}|^{-1/2} \sigma^2 \int \left(n|\mathbf{H}| \int \int K(\mathbf{p}) K(\mathbf{q}) \rho_n(\mathbf{H}(\mathbf{p} - \mathbf{q})) d\mathbf{p} d\mathbf{q} + o(1) \right) \\
\cdot \;& w(\mathbf{x}) d\mathbf{x}.
\end{aligned}$$

Under the assumptions (A4), (A7), (A8) and (A9), as shown in Liu (2001),

$$\lim_{n\to\infty} n|\mathbf{H}| \int K(\mathbf{p})K(\mathbf{q})\rho_n(\mathbf{H}(\mathbf{p}-\mathbf{q}))d\mathbf{p}d\mathbf{q} = K^{(2)}(0)\rho_c,$$

and, therefore,

$$\mathbb{E}\left(I_{32}\right) = |\mathbf{H}|^{-1/2}\sigma^2 K^{(2)}(0)\rho_c \int w(\mathbf{x})d\mathbf{x}\cdot\{1+o(1)\}, \tag{13}$$

corresponding to the second term of $b_{0\mathbf{H}}$ in Theorem 1 of the main paper.

The variance of I_{32} can be computed considering that:

$$\mathrm{Var}(I_{32}) = \mathbb{E}(\mathrm{Var}(I_{32}|\mathbf{X}_1,\ldots,\mathbf{X}_n)) + \mathrm{Var}(\mathbb{E}(I_{32}|\mathbf{X}_1,\ldots,\mathbf{X}_n)). \tag{14}$$

Let

$$W_{ij} = \int \frac{1}{f^2(\mathbf{x})} K_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{x}\right) K_{\mathbf{H}}\left(\mathbf{X}_j-\mathbf{x}\right) w(\mathbf{x})d\mathbf{x},$$

thus,

$$\begin{aligned}
\mathrm{Var}(I_{32}|\mathbf{X}_1,\ldots,\mathbf{X}_n) &= \mathrm{Var}\left(n^{-1}|\mathbf{H}|^{1/2}\sum_{i\neq j} W_{ij}\varepsilon_i\varepsilon_j|\mathbf{X}_1,\ldots,\mathbf{X}_n\right)\\
&= 4n^{-2}|\mathbf{H}|\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{k=1}^{n-1}\sum_{l=k+1}^{n} W_{ij}W_{kl}\mathrm{Cov}(\varepsilon_i\varepsilon_j,\varepsilon_k\varepsilon_l)\\
&= T_{31}+T_{32}+T_{33},
\end{aligned} \tag{15}$$

where

$$\begin{aligned}
T_{31} &= 4n^{-2}|\mathbf{H}|\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} W_{ij}^2\mathrm{Cov}(\varepsilon_i\varepsilon_j,\varepsilon_i\varepsilon_j),\\
T_{32} &= 4n^{-2}|\mathbf{H}|\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{l=i+2}^{n} W_{ij}W_{il}\mathrm{Cov}(\varepsilon_i\varepsilon_j,\varepsilon_i\varepsilon_l),\\
T_{33} &= 4n^{-2}|\mathbf{H}|\sum_{\text{all different indices } i,\, j,\, k,\, l} W_{ij}W_{kl}\mathrm{Cov}(\varepsilon_i\varepsilon_j,\varepsilon_k\varepsilon_l).
\end{aligned}$$

First, when $i=k$ and $j=l$, the total number of terms is $n(n-1)/2$. Second, when one of the i and j is equal to one of the k and l (without loss of generality, assume $i=k$ and $j\neq l$), the total number of terms can be bounded by n^3. Finally, when $i,j,k,$ and l are all different, the total number of terms can be bounded by n^4.

The expected value of $\mathrm{Var}(I_{32}|\mathbf{X}_1,\ldots,\mathbf{X}_n)$ is computed, calculating the mean of the terms T_{31}, T_{32}, and T_{33},

$$\mathbb{E}(\mathrm{Var}(I_{32}|\mathbf{X}_1,\ldots,\mathbf{X}_n)) = \mathbb{E}(T_{31})+\mathbb{E}(T_{32})+\mathbb{E}(T_{33}). \tag{16}$$

As for T_{31}, using assumption (A5), this term can be split as

$$\begin{aligned}
T_{31} &= 4n^{-2}|\mathbf{H}|\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} W_{ij}^2\mathrm{Cov}(\varepsilon_i\varepsilon_j,\varepsilon_i\varepsilon_j)\\
&= 4n^{-2}|\mathbf{H}|\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} W_{ij}^2[\sigma^4+\mathrm{Cov}^2(\varepsilon_i,\varepsilon_j)]\\
&= T_{311}+T_{312},
\end{aligned}$$

where

$$T_{311} = 4\sigma^4 n^{-2}|\mathbf{H}| \int\int \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{t}) K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{t}) \cdot w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t},$$

and

$$T_{312} = 4\sigma^4 n^{-2}|\mathbf{H}| \int\int \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{t}) K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{t}) \cdot \rho_n^2(\mathbf{X}_i - \mathbf{X}_j) w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t}.$$

Taking into account the first part of Lemma 2,

$$\begin{aligned}
\mathbb{E}(T_{311}) &= 4\sigma^4|\mathbf{H}|\frac{n-1}{2n} \int\int \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} \left[|\mathbf{H}|^{-1}(K^{(2)}(\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t}))\{f(\mathbf{t}) + o(1)\}\right]^2 w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t} \\
&= 2\sigma^4|\mathbf{H}|\frac{n-1}{n} \int\int \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} |\mathbf{H}|^{-2}(K^{(2)}(\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t}))^2 f^2(\mathbf{t}) w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t} \cdot \{1 + o(1)\} \\
&= 2\sigma^4\frac{n-1}{n} \int\int \frac{1}{f^2(\mathbf{x})}(K^{(2)}(\mathbf{p}))^2 w(\mathbf{x})w(\mathbf{x}+\mathbf{H}\mathbf{p})d\mathbf{x}d\mathbf{t} \cdot \{1 + o(1)\} \\
&= 2\frac{n-1}{n}\sigma^4 K^{(4)}(0) \int \frac{w^2(\mathbf{x})}{f^2(\mathbf{x})} d\mathbf{x} \cdot \{1 + o(1)\} \\
&= 2\sigma^4 K^{(4)}(0) \int \frac{w^2(\mathbf{x})}{f^2(\mathbf{x})} d\mathbf{x} \cdot \{1 + o(1)\}. \qquad (17)
\end{aligned}$$

Similarly for T_{312}, using assumptions (A2), (A3) and (A7), and taking into account Lemma 5, this term becomes

$$\begin{aligned}
\mathbb{E}(T_{312}) &= 4\sigma^4|\mathbf{H}|\frac{n-1}{2n} \int\int \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} \left[|\mathbf{H}|^{-2}\left(f^2(\mathbf{t}) \int\int K(-\mathbf{p} + \mathbf{H}^{-1}(\mathbf{x}-\mathbf{t}))\right.\right. \\
&\quad \cdot \left.\left. K(-\mathbf{q} + \mathbf{H}^{-1}(\mathbf{x}-\mathbf{t}))K(\mathbf{p})K(\mathbf{q})\rho_n^2(\mathbf{H}(\mathbf{p}-\mathbf{q}))d\mathbf{p}d\mathbf{q} \cdot \{1 + o(1)\}\right)\right] w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t} \\
&= 2\frac{n-1}{n}\sigma^4 \int\int\int\int \frac{1}{f^2(\mathbf{x})} K(-\mathbf{p}+\mathbf{u}) K(-\mathbf{q}+\mathbf{u}) K(\mathbf{p}) K(\mathbf{q}) w(\mathbf{x})w(\mathbf{x}+\mathbf{H}\mathbf{u}) \\
&\quad \cdot \rho_n^2(\mathbf{H}(\mathbf{p}-\mathbf{q}))d\mathbf{p}d\mathbf{q}d\mathbf{x}d\mathbf{u} \cdot \{1 + o(1)\} \\
&\leq 2\frac{n-1}{n^2|\mathbf{H}|}\frac{\sigma^4 K_M^4 w_M^2}{f_M^2} \int \{n|\mathbf{H}| \int \rho_n^2(\mathbf{H}(\mathbf{p}-\mathbf{q}))d\mathbf{p}\}d\mathbf{q} \cdot \{1 + o(1)\},
\end{aligned}$$

where f_M denotes the lower bound of f (assumption (A3)).

Since

$$n|\mathbf{H}| \int \rho_n^2(\mathbf{H}(\mathbf{p}-\mathbf{q}))d\mathbf{p} \leq n \int |\rho_n(\mathbf{t})|d\mathbf{t} \leq C_1,$$

it is obtained that

$$\begin{aligned}
\mathbb{E}(T_{312}) &\leq 2\frac{\sigma^4 K_M^4 w_M^2}{f_M^2}\frac{C_1}{n|\mathbf{H}|}\frac{n-1}{n} \\
&= O_p(n^{-1}|\mathbf{H}|^{-1}). \qquad (18)
\end{aligned}$$

Then, from (17) and (18), it follows that

$$\mathbb{E}(T_{31}) = 2\sigma^4 K^{(4)}(0)\int \frac{w^2(\mathbf{x})}{f^2(\mathbf{x})}d\mathbf{x}\cdot\{1+o(1)\} + O_p(n^{-1}|\mathbf{H}|^{-1}). \tag{19}$$

With respect to the term T_{32} (corresponding to the case with $i = k$ and $j \neq l$ in (15)), using assumption (A5), it follows that

$$\begin{aligned}
T_{32} &= 4n^{-2}|\mathbf{H}|\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{j=i+2}^{n} W_{ij}W_{il}\mathrm{Cov}(\varepsilon_i\varepsilon_j,\varepsilon_i\varepsilon_l)\\
&= 4n^{-2}|\mathbf{H}|\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{j=i+2}^{n} W_{ij}W_{il}[\mathrm{Var}(\varepsilon_i)\mathrm{Cov}(\varepsilon_j,\varepsilon_l)+\mathrm{Cov}(\varepsilon_i,\varepsilon_l)\mathrm{Cov}(\varepsilon_j,\varepsilon_i)]\\
&= T_{321}+T_{322},
\end{aligned}$$

where

$$\begin{aligned}
T_{321} &= 4\sigma^4 n^{-2}|\mathbf{H}|\int\int\frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})}\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{l=i+2}^{n} K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{t})K_{\mathbf{H}}(\mathbf{X}_l-\mathbf{t})\\
&\cdot\ \rho_n(\mathbf{X}_j-\mathbf{X}_l)w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t},
\end{aligned}$$

and

$$\begin{aligned}
T_{322} &= 4\sigma^4 n^{-2}|\mathbf{H}|\int\int\frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})}\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{l=i+2}^{n} K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{t})K_{\mathbf{H}}(\mathbf{X}_l-\mathbf{t})\\
&\cdot\ \rho_n(\mathbf{X}_i-\mathbf{X}_l)\rho_n(\mathbf{X}_j-\mathbf{X}_i)w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t}.
\end{aligned}$$

Using the assumption (A4) and the first part of Lemma 2 and of Lemma 4, one gets

$$\begin{aligned}
\mathbb{E}(T_{321}) &= 4\sigma^4 n|\mathbf{H}|\int\int\frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})}\frac{1}{|\mathbf{H}|}(K^{(2)}(\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t}))f(\mathbf{t})\cdot\{1+o(1)\}\\
&\cdot\ \left(\int\int K(\mathbf{p})K(\mathbf{q})\rho_n(\mathbf{x}-\mathbf{t}+\mathbf{H}(\mathbf{p}-\mathbf{q}))d\mathbf{p}d\mathbf{q}f(\mathbf{x})f(\mathbf{t})\cdot\{1+o(1)\}\right)w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t}\\
&= 4\sigma^4 n\int\int\frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})}\left(K^{(2)}(\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t}))f(\mathbf{t})\int\int K(\mathbf{p})K(\mathbf{q})\rho_n(\mathbf{x}-\mathbf{t}+\mathbf{H}(\mathbf{p}-\mathbf{q}))\right.\\
&\quad \left. d\mathbf{p}d\mathbf{q}f(\mathbf{x})f(\mathbf{t})\right)w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t}\cdot\{1+o(1)\}\\
&= 4\sigma^4 n\int\int\frac{1}{f(\mathbf{x})}K^{(2)}(\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t}))\int\int K(\mathbf{p})K(\mathbf{q})\rho_n(\mathbf{x}-\mathbf{t}+\mathbf{H}(\mathbf{p}-\mathbf{q}))d\mathbf{p}d\mathbf{q}\\
&\cdot\ w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t}\{1+o(1)\}\\
&= 4\sigma^4 n|\mathbf{H}|\int\int\int\int\frac{1}{f(\mathbf{x})}K^{(2)}(\mathbf{r})K(\mathbf{p})K(\mathbf{q})w(\mathbf{x})w(\mathbf{x}-\mathbf{H}\mathbf{r})\rho_n(\mathbf{H}(\mathbf{p}-\mathbf{q}+\mathbf{r}))d\mathbf{p}d\mathbf{q}d\mathbf{x}d\mathbf{r}\\
&\cdot\ \{1+o(1)\}\\
&= 4\sigma^4\int\int\int\frac{1}{f(\mathbf{x})}K^{(2)}(\mathbf{r})K(\mathbf{q})w^2(\mathbf{x})\{n|\mathbf{H}|\int K(\mathbf{p})\rho_n(\mathbf{H}(\mathbf{p}-\mathbf{q}+\mathbf{r}))d\mathbf{p}\}d\mathbf{q}d\mathbf{x}d\mathbf{r}\\
&\cdot\ \{1+o(1)\}.
\end{aligned}$$

As it was shown in Liu (2001),

$$\lim_{n\to\infty} n|\mathbf{H}|\int K(\mathbf{p})\rho_n(\mathbf{H}(\mathbf{p}-\mathbf{q}+\mathbf{r}))d\mathbf{p} = K(\mathbf{q}-\mathbf{r})\rho_c,$$

and, therefore,

$$\begin{aligned}
\mathbb{E}(T_{321}) &= 4\sigma^4\rho_c \int\int\int \frac{1}{f(\mathbf{x})} K^{(2)}(\mathbf{r})\, K(\mathbf{q})\, K(\mathbf{r}-\mathbf{q}) w^2(\mathbf{x}) d\mathbf{q} d\mathbf{x} d\mathbf{r} \cdot \{1+o(1)\} \\
&= 4\sigma^4\rho_c \int\int \frac{1}{f(\mathbf{x})} (K^{(2)}(\mathbf{r}))^2 w^2(\mathbf{x}) d\mathbf{r} d\mathbf{x} \cdot \{1+o(1)\} \\
&= 4\sigma^4 K^{(4)}(0)\rho_c \int \frac{w^2(\mathbf{x})}{f(\mathbf{x})} d\mathbf{x} \cdot \{1+o(1)\}. \qquad (20)
\end{aligned}$$

Similarly, taking into account that K is bounded, assumption (A4) and Lemma 6, the expected value of T_{322} becomes

$$\begin{aligned}
\mathbb{E}(T_{322}) &= 4\sigma^4 n|\mathbf{H}| \int\int \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} |\mathbf{H}|^{-1} \Big(f^2(\mathbf{x}) f(\mathbf{t}) \int\int\int K(\mathbf{p}) K(\mathbf{q})\, K\left(-\mathbf{p}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})\right) K(\mathbf{r}) \\
&\cdot \quad \rho_n(\mathbf{H}(\mathbf{p}-\mathbf{q}))\rho_n(\mathbf{x}-\mathbf{t}+\mathbf{H}(\mathbf{p}-\mathbf{r})) d\mathbf{p} d\mathbf{q} d\mathbf{r} \cdot \{1+o(1)\} \Big) w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \\
&= 4\sigma^4 n \int\int\int\int\int \frac{1}{f(\mathbf{t})} K(\mathbf{p})\, K\left(-\mathbf{p}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})\right) K(\mathbf{q})\, K(\mathbf{r})\, w(\mathbf{x}) w(\mathbf{t}) \\
&\cdot \quad \rho_n(\mathbf{x}-\mathbf{t}+\mathbf{H}(\mathbf{p}-\mathbf{r}))\rho_n(\mathbf{H}(\mathbf{p}-\mathbf{q})) d\mathbf{p} d\mathbf{q} d\mathbf{r} d\mathbf{x} d\mathbf{t} \cdot \{1+o(1)\} \\
&= 4\sigma^4 n|\mathbf{H}| \int\int\int\int\int \frac{1}{f(\mathbf{t})} K(\mathbf{p})\, K(-\mathbf{p}+\mathbf{u}))\, K(\mathbf{q})\, K(\mathbf{r})\, w(\mathbf{t}+\mathbf{H}\mathbf{u}) w(\mathbf{t}) \\
&\cdot \quad \rho_n(\mathbf{H}(\mathbf{p}-\mathbf{r}+\mathbf{u}))\rho_n(\mathbf{H}(\mathbf{p}-\mathbf{q})) d\mathbf{p} d\mathbf{q} d\mathbf{r} d\mathbf{u} d\mathbf{t} \cdot \{1+o(1)\} \\
&= 4\sigma^4 n^{-1}|\mathbf{H}|^{-1} \int\int\int \frac{1}{f(\mathbf{t})} K(\mathbf{p})\, K(-\mathbf{p}+\mathbf{u})\, w^2(\mathbf{t}) \{n|\mathbf{H}| \int K(\mathbf{r})\, \rho_n(\mathbf{H}(\mathbf{p}-\mathbf{r}+\mathbf{u})) d\mathbf{r}\} \\
&\cdot \quad \{n|\mathbf{H}| \int K(\mathbf{q})\, \rho_n(\mathbf{H}(\mathbf{p}-\mathbf{q})) d\mathbf{q}\} d\mathbf{p} d\mathbf{u} d\mathbf{t} \cdot \{1+o(1)\}.
\end{aligned}$$

Since

$$\lim_{n\to\infty} n|\mathbf{H}| \int K(\mathbf{r})\rho_n(\mathbf{H}(\mathbf{p}-\mathbf{r}+\mathbf{u})) d\mathbf{r} = K(\mathbf{p}+\mathbf{u})\rho_c,$$

$$\lim_{n\to\infty} n|\mathbf{H}| \int K(\mathbf{q})\rho_n(\mathbf{H}(\mathbf{p}-\mathbf{q})) d\mathbf{q} = K(\mathbf{p})\rho_c,$$

and taking into account that the functions K, w are bounded, and f is bounded away from zero, it follows that

$$\mathbb{E}(T_{322}) = O_p(n^{-1}|\mathbf{H}|^{-1}). \qquad (21)$$

Then, from (20) and (21), one gets that

$$\mathbb{E}(T_{32}) = 4\sigma^4 K^{(4)}(0)\rho_c \int \frac{w^2(\mathbf{x})}{f(\mathbf{x})} d\mathbf{x} \cdot \{1+o(1)\} + O_p(n^{-1}|\mathbf{H}|^{-1}). \qquad (22)$$

Regarding the term T_{33} (when all i, j, k, l are different in (15)), using assumption (A5), it follows that

$$\begin{aligned}
T_{33} &= 4n^{-2}|\mathbf{H}| \sum_{\text{all different indices } i,\ j,\ k,\ l} W_{ij}W_{kl}\mathrm{Cov}(\varepsilon_i\varepsilon_j, \varepsilon_k\varepsilon_l) \\
&= 4n^{-2}|\mathbf{H}| \sum_{\text{all different indices } i,\ j,\ k,\ l} W_{ij}W_{kl}[\mathrm{Cov}(\varepsilon_i, \varepsilon_k)\mathrm{Cov}(\varepsilon_j, \varepsilon_l) \\
&+ \mathrm{Cov}(\varepsilon_i, \varepsilon_l)\mathrm{Cov}(\varepsilon_j, \varepsilon_k)] \\
&= T_{331} + T_{332},
\end{aligned}$$

where

$$\begin{aligned} T_{331} &= 4\sigma^4 n^{-2}|\mathbf{H}| \sum_{\text{all different indices } i,\, j,\, k,\, l} \int\int \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x}) \\ &\cdot \quad K_{\mathbf{H}}(\mathbf{X}_k - \mathbf{t}) K_{\mathbf{H}}(\mathbf{X}_l - \mathbf{t}) w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \rho_n(\mathbf{X}_i - \mathbf{X}_k)\rho_n(\mathbf{X}_j - \mathbf{X}_l), \end{aligned}$$

and

$$\begin{aligned} T_{332} &= 4\sigma^4 n^{-2}|\mathbf{H}| \sum_{\text{all different indices } i,\, j,\, k,\, l} \int\int \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x}) \\ &\cdot \quad K_{\mathbf{H}}(\mathbf{X}_k - \mathbf{t}) K_{\mathbf{H}}(\mathbf{X}_l - \mathbf{t}) w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \rho_n(\mathbf{X}_i - \mathbf{X}_l)\rho_n(\mathbf{X}_j - \mathbf{X}_k). \end{aligned}$$

Using the assumption (A4) and Lemma 4,

$$\begin{aligned} \mathbb{E}(T_{331}) &= 4\sigma^4 n^2 |\mathbf{H}| \int\int \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} \Big[f(\mathbf{x})f(\mathbf{t}) \int\int K(\mathbf{p}) K(\mathbf{q}) \rho_n(\mathbf{x} - \mathbf{t} + \mathbf{H}(\mathbf{p}-\mathbf{q})) d\mathbf{p} d\mathbf{q} \\ &\cdot \quad \{1 + o(1)\}\Big]^2 w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \\ &= 4\sigma^4 n^2 |\mathbf{H}| \int\int \Big[\int\int K(\mathbf{p}) K(\mathbf{q}) \rho_n(\mathbf{x} - \mathbf{t} + \mathbf{H}(\mathbf{p}-\mathbf{q})) d\mathbf{p} d\mathbf{q} \\ &\cdot \quad \int\int K(\mathbf{m}) K(\mathbf{r}) \rho_n(\mathbf{x} - \mathbf{t} + \mathbf{H}(\mathbf{m}-\mathbf{r})) d\mathbf{m} d\mathbf{r} \Big] w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \cdot \{1 + o(1)\} \\ &= 4\sigma^4 n^2 |\mathbf{H}| \int\int\int\int\int\int K(\mathbf{p}) K(\mathbf{q}) K(\mathbf{m}) K(\mathbf{r}) w(\mathbf{x}) w(\mathbf{t}) \\ &\cdot \quad \rho_n(\mathbf{x} - \mathbf{t} + \mathbf{H}(\mathbf{p}-\mathbf{q}))\rho_n(\mathbf{x} - \mathbf{t} + \mathbf{H}(\mathbf{m}-\mathbf{r})) d\mathbf{p} d\mathbf{q} d\mathbf{m} d\mathbf{r} d\mathbf{x} d\mathbf{t} \cdot \{1 + o(1)\} \\ &= 4\sigma^4 n^2 |\mathbf{H}|^2 \int\int\int\int\int\int K(\mathbf{p}) K(\mathbf{q}) K(\mathbf{m}) K(\mathbf{r}) w(\mathbf{x}) w(\mathbf{x} - \mathbf{H}\mathbf{u}) \\ &\cdot \quad \rho_n(\mathbf{H}(\mathbf{p}-\mathbf{q}+\mathbf{u}))\rho_n(\mathbf{H}(\mathbf{m}-\mathbf{r}+\mathbf{u})) d\mathbf{p} d\mathbf{q} d\mathbf{m} d\mathbf{r} d\mathbf{x} d\mathbf{u} \cdot \{1 + o(1)\} \\ &= 4\sigma^4 \int\int\int\int K(\mathbf{q}) K(\mathbf{r}) w^2(\mathbf{x}) \{n|\mathbf{H}| \int K(\mathbf{m}) \rho_n(\mathbf{H}(\mathbf{m}-\mathbf{r}+\mathbf{u})) d\mathbf{m}\} \\ &\cdot \quad \{n|\mathbf{H}| \int K(\mathbf{p}) \rho_n(\mathbf{H}(\mathbf{p}-\mathbf{q}+\mathbf{u})) d\mathbf{p}\} d\mathbf{q} d\mathbf{r} d\mathbf{x} d\mathbf{u} \cdot \{1 + o(1)\}. \end{aligned}$$

Since

$$\lim_{n\to\infty} n|\mathbf{H}| \int K(\mathbf{p})\rho_n(\mathbf{H}(\mathbf{p}-\mathbf{q}+\mathbf{u}))d\mathbf{p} = K(\mathbf{q}-\mathbf{u})\rho_c,$$

and

$$\lim_{n\to\infty} n|\mathbf{H}| \int K(\mathbf{m})\rho_n(\mathbf{H}(\mathbf{m}-\mathbf{r}+\mathbf{u}))d\mathbf{m} = K(\mathbf{r}-\mathbf{u})\rho_c,$$

it follows that

$$\begin{aligned} \mathbb{E}(T_{331}) &= 4\sigma^4 \rho_c^2 \int\int\int\int K(\mathbf{q}) K(\mathbf{u}-\mathbf{q}) K(\mathbf{r}) K(\mathbf{u}-\mathbf{r}) w^2(\mathbf{x}) d\mathbf{q} d\mathbf{r} d\mathbf{x} d\mathbf{u} \cdot \{1 + o(1)\} \\ &= 4\sigma^4 \rho_c^2 \int\int (K^{(2)}(\mathbf{u}))^2 w^2(\mathbf{x}) d\mathbf{x} d\mathbf{u} \cdot \{1 + o(1)\} \\ &= 4\sigma^4 \rho_c^2 K^{(4)}(0) \int w^2(\mathbf{x}) d\mathbf{x} \cdot \{1 + o(1)\}. \end{aligned} \tag{23}$$

For symmetry, $\mathbb{E}(T_{332}) = \mathbb{E}(T_{331})$ and, therefore, using (23), it follows that

$$\mathbb{E}(T_{33}) = 8\sigma^4 K^{(4)}(0)\rho_c^2 \int w^2(\mathbf{x}) d\mathbf{x} \cdot \{1 + o(1)\}. \tag{24}$$

So, from (16), (19), (22) and (24), it is obtained that

$$
\begin{aligned}
\mathbb{E}(\mathrm{Var}(I_{32}|\mathbf{X}_1,\ldots,\mathbf{X}_n)) &= 2\sigma^4 K^{(4)}(0)\int \frac{w^2(\mathbf{x})}{f^2(\mathbf{x})}d\mathbf{x}\cdot\{1+o(1)\}+O_p(n^{-1}|\mathbf{H}|^{-1}) \\
&+ 4\sigma^4 K^{(4)}(0)\rho_c\int \frac{w^2(\mathbf{x})}{f(\mathbf{x})}d\mathbf{x}\cdot\{1+o(1)\}+O_p(n^{-1}|\mathbf{H}|^{-1}) \\
&+ 8\sigma^4 K^{(4)}(0)\rho_c^2\int w^2(\mathbf{x})d\mathbf{x}\cdot\{1+o(1)\}. \qquad (25)
\end{aligned}
$$

With respect to the $\mathrm{Var}(\mathbb{E}(I_{32}|\mathbf{X}_1,\ldots,\mathbf{X}_n))$, the second term in equation (14), denoting by

$$
\phi_{ij}=\int \frac{1}{f^2(\mathbf{x})}K_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{x}\right)K_{\mathbf{H}}\left(\mathbf{X}_j-\mathbf{x}\right)\rho_n(\mathbf{X}_i-\mathbf{X}_j)w(\mathbf{x})d\mathbf{x},
$$

and using the expression of the $\mathbb{E}(I_{32}|\mathbf{X}_1,\ldots,\mathbf{X}_n)$, given in (12), it can be split as:

$$
\begin{aligned}
\mathrm{Var}(\mathbb{E}(I_{32}|\mathbf{X}_1,\ldots,\mathbf{X}_n)) &= \mathrm{Var}\left(|\mathbf{H}|^{1/2}\sigma^2\frac{1}{n}\sum_{i\neq j}\phi_{ij}\right) \\
&= 4\sigma^4 n^{-2}|\mathbf{H}|\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{k=1}^{n-1}\sum_{l=k+1}^{n}\mathrm{Cov}(\phi_{ij},\phi_{kl}). \qquad (26)
\end{aligned}
$$

Now, consider the value of $\mathrm{Cov}(\phi_{ij},\phi_{kl})$ according to the following three exclusive cases. First, when $i=k$ and $j=l$, the total number of such terms is $n(n-1)/2$. In this case, using Lemma 5, one gets

$$
\begin{aligned}
\mathrm{Cov}(\phi_{ij},\phi_{ij}) &\leq \mathbb{E}\bigg(\int\int \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})}K_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{x}\right)K_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{t}\right)K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})K_{\mathbf{H}}\left(\mathbf{X}_j-\mathbf{t}\right) \\
&\cdot\ \rho_n^2(\mathbf{X}_i-\mathbf{X}_j)w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t}\bigg) \\
&= \int\int \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})}|\mathbf{H}|^{-2}\bigg(f^2(\mathbf{t})\int\int K\left(-\mathbf{p}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})\right)K\left(-\mathbf{q}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})\right) \\
&\cdot\ K\left(\mathbf{p}\right)K\left(\mathbf{q}\right)\rho_n^2(\mathbf{H}(\mathbf{p}-\mathbf{q}))d\mathbf{p}d\mathbf{q}+o(1)\bigg)w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t} \\
&= |\mathbf{H}|^{-1}\int\int\int\int \frac{1}{f^2(\mathbf{x})}K\left(-\mathbf{p}+\mathbf{u}\right)K\left(-\mathbf{q}+\mathbf{u}\right)K\left(\mathbf{p}\right)K\left(\mathbf{q}\right)w(\mathbf{x})w(\mathbf{x}-\mathbf{H}\mathbf{u}) \\
&\cdot\ \rho_n^2(\mathbf{H}(\mathbf{p}-\mathbf{q}))d\mathbf{p}d\mathbf{q}d\mathbf{x}d\mathbf{u}\cdot\{1+o(1)\} \\
&\leq \frac{K_M^4 w_M^2}{f_M^2 n|\mathbf{H}|^2}\int\{n|\mathbf{H}|\int\rho_n^2(\mathbf{H}(\mathbf{p}-\mathbf{q}))d\mathbf{p}\}d\mathbf{q}\cdot\{1+o(1)\}.
\end{aligned}
$$

Since

$$
n|\mathbf{H}|\int\rho_n^2(\mathbf{H}(\mathbf{p}-\mathbf{q}))d\mathbf{p}\leq n\int|\rho_n(\mathbf{t})|d\mathbf{t}\leq C_2,
$$

then

$$
\mathrm{Cov}(\phi_{ij},\phi_{ij}) \leq \frac{K_M^4 w_M^2}{f_M^2}\frac{C_2}{n|\mathbf{H}|^2}. \qquad (27)
$$

Second, when $i = k$ and $j \neq l$ in (26). In this case, the total number of such terms can be bounded by n^3. Using Lemma 6, it follows that

$$\begin{aligned}
\mathrm{Cov}(\phi_{ij}, \phi_{il}) &= \mathbb{E}(\phi_{ij}, \phi_{il}) - \mathbb{E}(\phi_{ij})\mathbb{E}(\phi_{il}) \\
&= \mathbb{E}(\phi_{ij}, \phi_{il}) - (\mathbb{E}(\phi_{ij}))^2 \\
&\leq \mathbb{E}(\phi_{ij}, \phi_{il}) \\
&= \mathbb{E}\Big(\int\int \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{t}) K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_l - \mathbf{t}) \\
&\cdot \quad \rho_n(\mathbf{X}_i - \mathbf{X}_j)\rho_n(\mathbf{X}_i - \mathbf{X}_l) w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \Big) \\
&= \int\int \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} |\mathbf{H}|^{-1} \Big(f^2(\mathbf{x}) f(\mathbf{t}) \int\int\int K(\mathbf{p}) K(\mathbf{q}) K\left(-\mathbf{p} + \mathbf{H}^{-1}(\mathbf{x} - \mathbf{t})\right) K(\mathbf{r}) \\
&\cdot \quad \rho_n(\mathbf{H}(\mathbf{p} - \mathbf{q}))\rho_n(\mathbf{x} - \mathbf{t} + \mathbf{H}(\mathbf{p} - \mathbf{r})) d\mathbf{p} d\mathbf{q} d\mathbf{r} \cdot \{1 + o(1)\} \Big) w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \\
&= \int\int\int\int\int \frac{1}{f(\mathbf{t})} K(\mathbf{p}) K(-\mathbf{p} + \mathbf{u}) K(\mathbf{q}) K(\mathbf{r}) w(\mathbf{t} + \mathbf{H}\mathbf{u}) w(\mathbf{t}) \\
&\cdot \quad \rho_n(\mathbf{H}(\mathbf{p} - \mathbf{r} + \mathbf{u}))\rho_n(\mathbf{H}(\mathbf{p} - \mathbf{q})) d\mathbf{p} d\mathbf{q} d\mathbf{r} d\mathbf{u} d\mathbf{t} \cdot \{1 + o(1)\} \\
&= n^{-2}|\mathbf{H}|^{-2} \int\int\int \frac{1}{f(\mathbf{t})} K(\mathbf{p}) K(-\mathbf{p} + \mathbf{u}) w^2(\mathbf{t}) \{ n|\mathbf{H}| \int K(\mathbf{r}) \rho_n(\mathbf{H}(\mathbf{p} - \mathbf{r} + \mathbf{u})) d\mathbf{r} \} \\
&\cdot \quad \{ n|\mathbf{H}| \int K(\mathbf{q}) \rho_n(\mathbf{H}(\mathbf{p} - \mathbf{q})) d\mathbf{q} \} d\mathbf{p} d\mathbf{u} d\mathbf{t} \cdot \{1 + o(1)\}.
\end{aligned}$$

Since

$$\lim_{n \to \infty} n|\mathbf{H}| \int K(\mathbf{r})\rho_n(\mathbf{H}(\mathbf{p} - \mathbf{r} + \mathbf{u})) d\mathbf{r} = K(\mathbf{p} + \mathbf{u})\rho_c,$$

$$\lim_{n \to \infty} n|\mathbf{H}| \int K(\mathbf{q})\rho_n(\mathbf{H}(\mathbf{p} - \mathbf{q})) d\mathbf{q} = K(\mathbf{p})\rho_c,$$

and taking into account that the functions K, w are bounded, and f is bounded away from zero, it is obtained that

$$\mathrm{Cov}(\phi_{ij}, \phi_{il}) \leq \frac{C_3}{n^2|\mathbf{H}|^2}. \tag{28}$$

Finally, when i, j, k, l are all distinct in (26), as ϕ_{ij} and ϕ_{kl} are independent,

$$\mathrm{Cov}(\phi_{ij}, \phi_{kl}) = 0, \tag{29}$$

Then, considering (26), (27), (28) and (29), it follows that

$$\begin{aligned}
\mathrm{Var}(\mathbb{E}(I_{32}|\mathbf{X}_1, \ldots, \mathbf{X}_n)) &= \mathrm{Var}\left(|\mathbf{H}|^{1/2}\sigma^2 \frac{1}{n} \sum_{i \neq j} \phi_{ij} \right) \\
&= 4\sigma^4 n^{-2} |\mathbf{H}| \left(\frac{n^2 - n}{2} \frac{C_2}{n|\mathbf{H}|^2} + n^3 \frac{C_3}{n^2|\mathbf{H}|^2} \right) \\
&= O_p(n^{-1}|\mathbf{H}|^{-1}).
\end{aligned} \tag{30}$$

Now, from (14), (25) and (30), the leading term of the variance of I_{32} is given by:

$$V = 2\sigma^4 K^{(4)}(0) \Big[\int \frac{w^2(\mathbf{x})}{f^2(\mathbf{x})} d\mathbf{x} + 2\rho_c \int \frac{w^2(\mathbf{x})}{f(\mathbf{x})} d\mathbf{x} + 4\rho_c^2 \int w^2(\mathbf{x}) d\mathbf{x} \Big]. \tag{31}$$

Therefore, using the central limit theorem for degenerate reduced U-statistics under $\alpha-$mixing conditions, given by Kim et al. (2013), it is obtained that the term I_{32} converges, in distribution, to a normal distribution with mean the leading term of (13) and variance given by (31).

On the other hand, in virtue of the Cauchy-Schwarz inequality, the cross terms in T_{n1} resulting from the products of $I_1(\mathbf{x})$, $I_2(\mathbf{x})$ and $I_3(\mathbf{x})$ are all of smaller order. Therefore, combining the results in (3), (4) and (11), and the asymptotic normality of I_{32} (with bias the leading term of (13) and variance (31)), one gets

$$V^{-1/2}(T_{n1} - b_{0\mathbf{H}} - b_{1\mathbf{H}}) \to_{\mathcal{L}} N(0,1) \text{ as } n \to \infty, \tag{32}$$

where

$$\begin{aligned} b_{0\mathbf{H}} &= |\mathbf{H}|^{-1/2}\sigma^2 K^{(2)}(\mathbf{0})\left[\int \frac{w(\mathbf{x})}{f(\mathbf{x})}d\mathbf{x} + \rho_c \int w(\mathbf{x})d\mathbf{x}\right], \\ b_{1\mathbf{H}} &= \int (K_{\mathbf{H}} * g(\mathbf{x}))^2 w(\mathbf{x})d\mathbf{x}, \end{aligned}$$

and

$$V = 2\sigma^4 K^{(4)}(0)\left[\int \frac{w^2(\mathbf{x})}{f^2(\mathbf{x})}d\mathbf{x} + 2\rho_c \int \frac{w^2(\mathbf{x})}{f(\mathbf{x})}d\mathbf{x} + 4\rho_c^2 \int w^2(\mathbf{x})d\mathbf{x}\right].$$

The term T_{n2} in T_n is of smaller order than T_{n1} (specifically, $T_{n2} = O_p(\text{tr}(\mathbf{H}^2)T_{n1})$), and by the Cauchy-Schwarz inequality, the cross term T_{n12} is of smaller order as well. Therefore, from (2), it follows that

$$T_n = T_{n1} + O_p(tr(\mathbf{H}^2)) + O_p(n^{-2/(4+d)+\eta}).$$

Taking into account (32), it follows that

$$V^{-1/2}(T_n - b_{0\mathbf{H}} - b_{1\mathbf{H}}) \to_{\mathcal{L}} N(0,1) \text{ as } n \to \infty,$$

with $b_{0\mathbf{H}}$, $b_{1\mathbf{H}}$ and V given above.

■

2 Simulation results

In this section, additional simulations complementing the study presented in Section 4 of the main paper are presented. This section is organized as follows. First, the asymptotic distribution of the test is illustrated with a particular example. The next subsections present an extension of the simulation results in the main paper, considering the use of non-scalar bandwidth matrices, employing a different regression function, assuming a random design, and including a nugget effect in dependence structure.

2.1 Asymptotic distribution of the test

Asymptotic distribution of test statistics are usually employed for test calibration in practice. However, the convergence of T_n to its limit distribution, as it happens with other smooth-based test, is too slow. This issue is pointed out in Section 3.2 of the main paper: the asymptotic distribution obtained in Theorem 1 could not be sufficiently precise when the sample size is small or medium. This was also noted in other nonparametric testing contexts (see, for example, Härdle and Mammen 1993). Moreover, the limit distribution of the test statistic depends on unknown quantities such as the design density and the error variance that, in a practical situation, must be estimated from the data. For these reasons, resampling methods are considered as an alternative to the asymptotic distribution. As shown in the main paper (and also in the following sections of this document), the bootstrap approach designed to be used in this context provides satisfactory results. Nevertheless, and for the sake of illustration, in this section, a brief simulation experiment is presented to study the performance of the asymptotic distribution of the test under the null hypothesis. Specifically, we consider the simple case of assuming f and σ^2 known, and the density estimator of $V^{-1/2}(T_n - b_{0\mathbf{H}})$ and the standard normal density function are compared.

A linear parametric regression family is chosen, $m_{\boldsymbol{\beta}}(X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$, being $\mathbf{X} = (X_1, X_2)$, and the regression function considered is:

$$m(X_1, X_2) = 2 + X_1 + X_2. \tag{33}$$

500 samples of sizes $n = 400$, 2500 and 10000 are generated from a regression model with explanatory variables drawn from a bivariate uniform distribution in the unit square, regression function (33), and random errors ε_i normally distributed with zero mean and with isotropic exponential covariance function:

$$\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = \sigma^2\{\exp(-\lambda n \|\mathbf{X}_i - \mathbf{X}_j\|)\}, \tag{34}$$

with values of $\sigma^2 = 0.4$ and $\lambda = 0.0005$. Note that with this selection λ, the values for the practical range are 5, 0.8 and 0.2, for $n = 400$, 2500 and 10000, respectively. The parametric fit was computed using the iterative least squares procedure described in Section 2.2 of the main paper, considering a linear model. The nonparametric fit was obtained using the multivariate local linear estimator with a multivariate Gaussian kernel and a scalar bandwidth matrix. With this kernel, the quantities $K^{(2)}(0)$ and $K^{(4)}(0)$ in the asymptotic bias and variance of T_n can be easily calculated. Additionally, considering (34), it is straightforward to prove that $\rho_c = 1/\lambda$. For simplicity, we also take $w(\mathbf{x}) = f(\mathbf{x})$, $\forall \mathbf{x} \in D \in \mathbb{R}^d$. For each sample and in every scenario, the statistic $V^{-1/2}(T_n - b_{0\mathbf{H}})$ is computed.

Figure 1 shows density estimates of $V^{-1/2}(T_n - b_{0\mathbf{H}})$ (blue lines), computed with a Gaussian kernel and the rule-of-thumb bandwidth selector, and the standard normal densities (red lines). The plot in the left panel corresponds to $n = 2500$ and the one in the right panel to $n = 10000$. When $n = 400$, the asymptotic distribution of $V^{-1/2}(T_n - b_{0\mathbf{H}})$ is very far from the standard normal distribution and it is not shown here. Only when the sample size is very large, the sampling distribution of the test statistic seems to approximate reasonably well the Gaussian limit distribution. It is expected that this approximation will be better for larger sample sizes. That means that to obtain reliable results with the asymptotic distribution of the test, it would be necessary to consider a huge sample size (ignoring f and σ^2, which should be estimated). In this situation, the application of the test will take an enormous computing time. In such scenarios, the use of binning techniques or big data methods could be of special interest to accelerate the running time when applying the test. These approaches are out of the scope of the present paper, but can be an interesting issue of research in future.

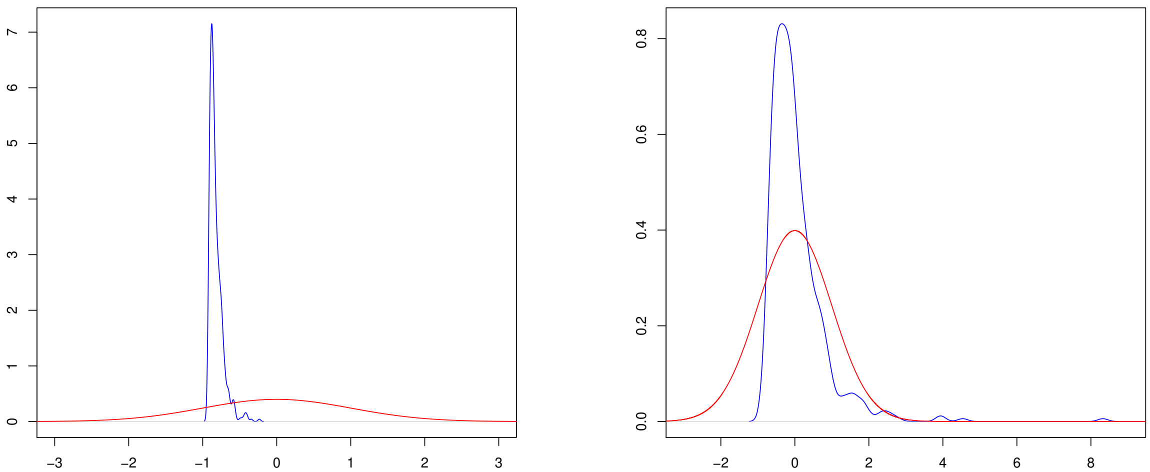

Figure 1: Density estimates of $V^{-1/2}(T_n - b_{0\mathbf{H}})$ (blue lines) and normal standard densities (red lines), considering $n = 2500$ (left panel) and $n = 10000$ (right panel).

2.2 Non-scalar bandwidths

This section contains additional simulations similar to those presented in the main paper, but taking a different type of bandwidth matrices to compute the nonparametric estimation of the regression function. While in the main paper, scalar matrix bandwidths (diagonal matrix with equal values in the main diagonal) were considered, here, diagonal bandwidths with different elements are used. A linear model $m_{\boldsymbol{\beta}}(X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ is chosen, and for different values of c (specifically, 0, 3 and 5) the regression function

$$m(X_1, X_2) = 2 + X_1 + X_2 + cX_1^3 \tag{35}$$

is considered. For each value of c, 500 samples of sizes $n = 225$ and 400 are generated on a bidimensional regular grid in the unit square, with regression function (35) and random errors ε_i normally distributed with zero mean and isotropic exponential covariance function:

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = \sigma^2\{\exp(-\|\mathbf{X}_i - \mathbf{X}_j\|/a_e)\}, \tag{36}$$

with $\sigma = 0.4$, 0.6, and 0.8. Different values of parameter a_e are considered: $a_e = 0.1$ (weak correlation), $a_e = 0.2$ (medium correlation) and $a_e = 0.4$ (strong correlation). No nugget effect is considered in this scenario.

Figure 2 shows the different exponential variogram models considered (brown lines for $\sigma = 0.4$, red lines for $\sigma = 0.6$, and orange lines for $\sigma = 0.8$. For each value of σ, solid, dashed and dotted lines for $a_e = 0.1, 0.2$ and 0.4, respectively).

Figure 3 shows, for $c = 0$, in the left panel, the regression function function (35) and, in the right panel, a simulated spatial process, considering $\sigma = 0.6$ and $a_e = 0.2$ in (36).

The regression functions, using (35), for $c = 3$ (left panel) and for $c = 5$ (right panel) are shown in Figure 4.

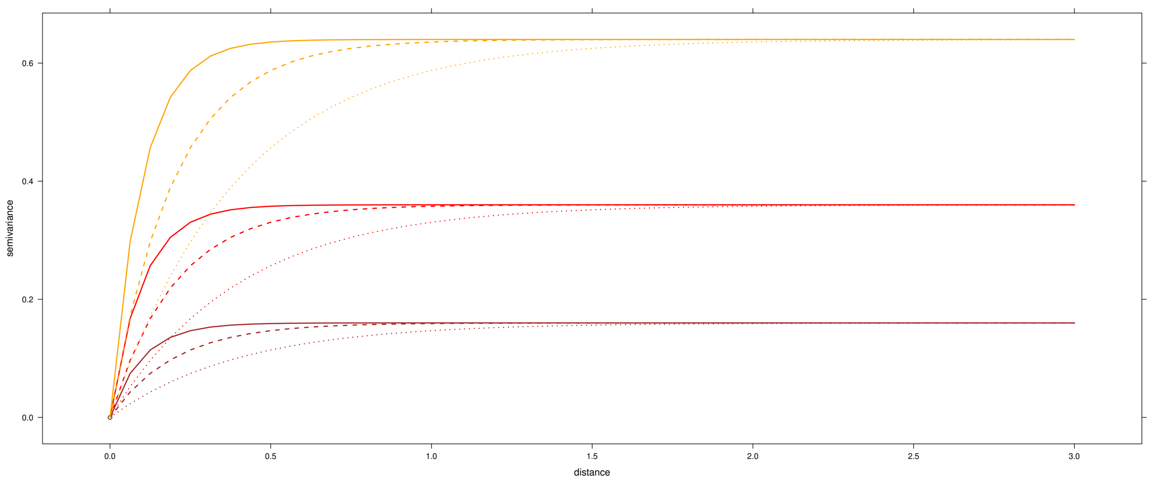

Figure 2: Exponential variogram models for the simulation scenario trying different bandwidth matrices.

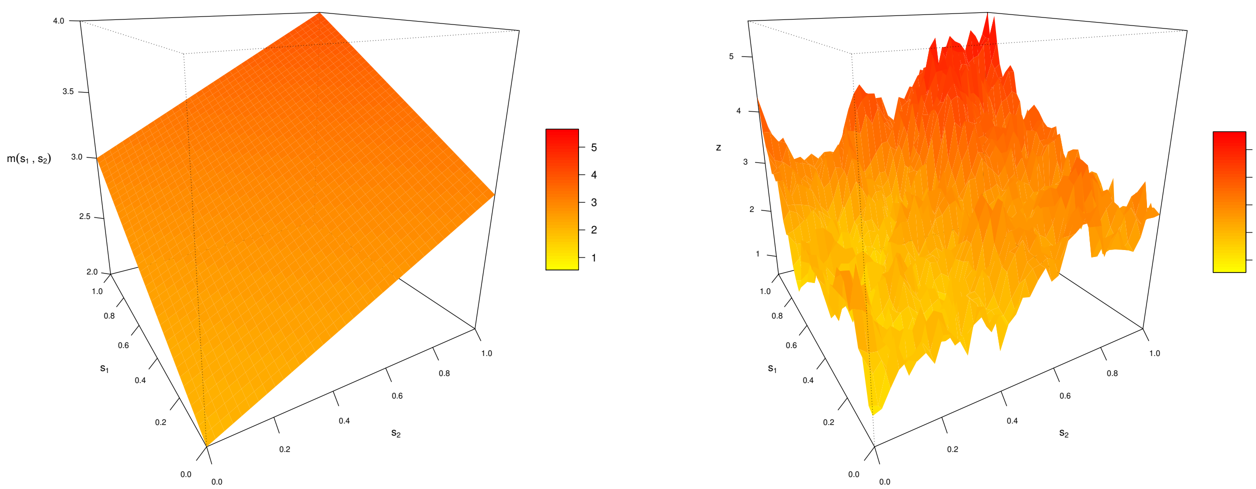

Figure 3: Regression model (35) for $c = 0$ (left panel) and a realization of the spatial process (right panel). The dependence structure of the errors is explained by an exponential covariogram with parameters $\sigma = 0.4$ and $a_e = 0.2$.

The bootstrap procedure described in Section 3.2 was applied, using $B = 500$ replicates. The weight function was taken constant with value 1. The parametric fit used for constructing (1) was computed using the iterative least squares procedure, considering a linear model, while the nonparametric fit was obtained using the multivariate local linear estimator estimator with a multiplicative triweight kernel. The bandwidth is taken as a diagonal matrix $\mathbf{H} = \text{diag}(h_1, h_2)$, being the values of h_1 and h_2 different.

Results are presented in Table 1, where the rejection proportions of the null hypothesis, for $\alpha = 0.05$, are displayed. Similarly to the results shown in the main paper, it can be observed that the test has a reasonable behavior. In particular, if $c = 0$ (under the null hypothesis) the rejection proportions are similar to the theoretical level, for the different values of h_1 and h_2 considered. For the alternative hypothesis ($c = 3$ and $c = 5$), the power of the test becomes larger as the value of c increases. On the other hand, the power of the test decreases with the point variance σ^2. In all scenarios, it can be seen that the rejection proportions depend on the bandwidth $\mathbf{H}$, especially, under the alternative

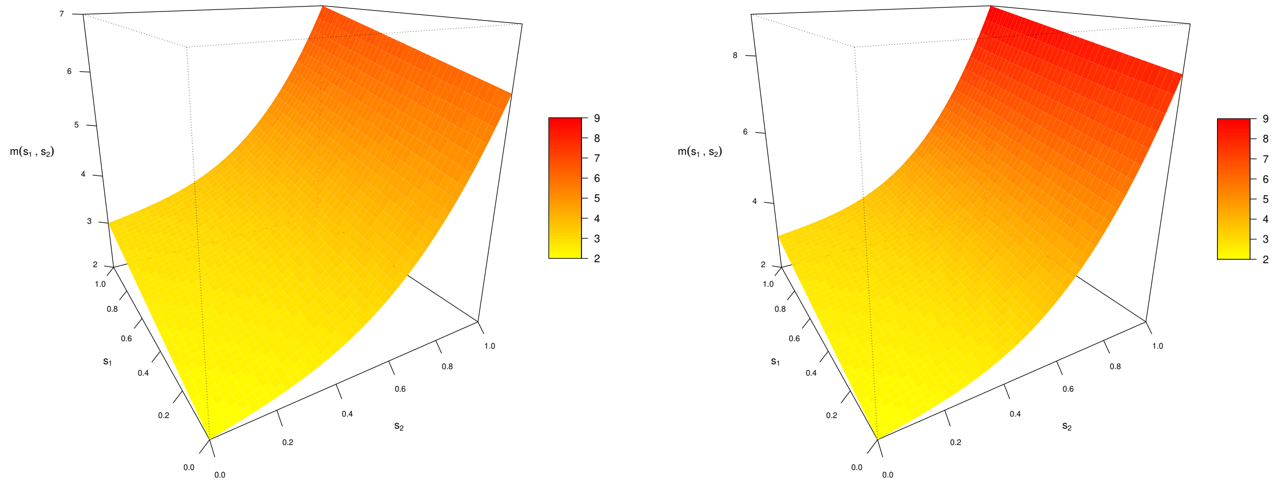

Figure 4: Regression model (35) for $c = 3$ (left panel) and $c = 5$ (right panel).

hypothesis.

For example, for a 15×15 grid, with $\sigma = 0.4$ and $a_e = 0.2$, it follows that, under the null hypothesis, the rejection proportions obtained are not significantly different from the theoretical level, considering both bandwidth matrices $\mathbf{H} = \text{diag}(1, 0.6)$ and $\mathbf{H} = \text{diag}(0.6, 1)$. However, the power of the test shows a different behavior. It is significantly larger when $\mathbf{H} = \text{diag}(0.6, 1)$ is considered. Then, under the alternative hypothesis, the rejection proportion depends on the values of h_1 and h_2. Note that, a comparison between Table 1 of the main paper and Table 1 of this supplementary material reveals that there are not relevant differences in terms of rejection proportions if $\mathbf{H} = \text{diag}(h, h)$ or $\mathbf{H} = \text{diag}(h_1, h_2)$ (with $h_1 \neq h_2$) is considered, for this particular scenario.

2.3 Alternative regression function

The second framework considered is similar to the previous regression scenario, but with mean function

$$m(X_1, X_2) = 3 + 2X_1 + X_2 + cx_1^3. \tag{37}$$

The errors of the model are also normally distributed with an exponential dependence structure, and the same parameters for c, σ, a_e, B, and n as in the previous framework are considered in this case. Table 2 shows the rejection proportions of the null hypothesis, for $\alpha = 0.05$, considering that the bandwidth is taken as a diagonal matrix $\mathbf{H} = \text{diag}(h, h)$, and different values of h are chosen, $h = 0.6, 0.7, 0.8, 0.9, 1$. Table 3 shows the results when the bandwidth is taken as a diagonal matrix $\mathbf{H} = \text{diag}(h_1, h_2)$, being the values of h_1 and h_2 different. It can be observed that considering different regression parameters ($\beta_0 = 3$, $\beta_1 = 2$ for the first coordinate and $\beta_2 = 1$ for the second one), the rejection proportions (under the null and the alternative hypothesis) are really similar to those obtained in the first setting (where $\beta_0 = 2$ and $\beta_1 = \beta_2 = 1$) and analogous conclusions can be deduced.

σ	a_e	c	n	**H**					
				$\begin{pmatrix}0.8 & 0\\ 0 & 0.6\end{pmatrix}$	$\begin{pmatrix}1 & 0\\ 0 & 0.6\end{pmatrix}$	$\begin{pmatrix}0.6 & 0\\ 0 & 0.8\end{pmatrix}$	$\begin{pmatrix}1 & 0\\ 0 & 0.8\end{pmatrix}$	$\begin{pmatrix}0.6 & 0\\ 0 & 1\end{pmatrix}$	$\begin{pmatrix}0.8 & 0\\ 0 & 1\end{pmatrix}$
0.4	0.1	0	225	0.074	0.054	0.066	0.038	0.052	0.036
			400	0.034	0.028	0.032	0.022	0.028	0.024
0.4	0.1	3	225	0.394	0.356	0.576	0.416	0.592	0.478
			400	0.298	0.236	0.502	0.322	0.530	0.404
0.4	0.1	5	225	0.998	0.994	1.000	0.998	1.000	1.000
			400	0.998	0.994	1.000	0.998	1.000	1.000
0.4	0.2	0	225	0.062	0.050	0.060	0.036	0.050	0.036
			400	0.054	0.038	0.054	0.024	0.038	0.022
0.4	0.2	3	225	0.780	0.726	0.870	0.786	0.876	0.822
			400	0.796	0.726	0.912	0.772	0.914	0.846
0.4	0.2	5	225	1.000	1.000	1.000	1.000	1.000	1.000
			400	0.998	1.000	1.000	1.000	1.000	1.000
0.4	0.4	0	225	0.126	0.106	0.132	0.070	0.092	0.080
			400	0.200	0.164	0.126	0.098	0.076	0.058
0.4	0.4	3	225	0.978	0.970	0.988	0.978	0.990	0.984
			400	0.980	0.974	0.992	0.984	0.992	0.988
0.4	0.4	5	225	1.000	1.000	1.000	1.000	1.000	1.000
			400	1.000	1.000	1.000	1.000	1.000	1.000
0.6	0.1	0	225	0.074	0.062	0.070	0.040	0.056	0.036
			400	0.034	0.028	0.032	0.022	0.028	0.022
0.6	0.1	3	225	0.060	0.050	0.154	0.062	0.166	0.086
			400	0.026	0.016	0.106	0.032	0.106	0.044
0.6	0.1	5	225	0.552	0.508	0.740	0.562	0.756	0.656
			400	0.476	0.434	0.700	0.494	0.708	0.570
0.6	0.2	0	225	0.062	0.050	0.062	0.036	0.050	0.036
			400	0.054	0.038	0.054	0.024	0.036	0.022
0.6	0.2	3	225	0.304	0.230	0.498	0.286	0.506	0.382
			400	0.322	0.268	0.502	0.314	0.504	0.380
0.6	0.2	5	225	0.888	0.860	0.944	0.882	0.942	0.924
			400	0.888	0.858	0.938	0.880	0.940	0.908
0.6	0.4	0	225	0.204	0.172	0.128	0.096	0.076	0.058
			400	0.198	0.164	0.126	0.100	0.076	0.058
0.6	0.4	3	225	0.734	0.658	0.844	0.700	0.838	0.752
			400	0.718	0.656	0.830	0.698	0.822	0.752
0.6	0.4	5	225	0.996	0.996	0.998	0.996	0.998	0.998
			400	0.994	0.994	0.996	0.994	0.996	0.994
0.8	0.1	0	225	0.072	0.060	0.068	0.038	0.052	0.036
			400	0.034	0.028	0.032	0.022	0.028	0.022
0.8	0.1	3	225	0.034	0.024	0.086	0.026	0.096	0.042
			400	0.008	0.006	0.054	0.004	0.052	0.010
0.8	0.1	5	225	0.154	0.132	0.328	0.180	0.360	0.230
			400	0.086	0.070	0.240	0.096	0.248	0.162
0.8	0.2	0	225	0.064	0.050	0.062	0.036	0.050	0.036
			400	0.052	0.038	0.050	0.024	0.036	0.022
0.8	0.2	3	225	0.144	0.112	0.280	0.122	0.268	0.166
			400	0.158	0.118	0.276	0.134	0.270	0.168
0.8	0.2	5	225	0.534	0.468	0.710	0.524	0.710	0.608
			400	0.556	0.472	0.722	0.546	0.722	0.628
0.8	0.4	0	225	0.126	0.110	0.134	0.068	0.094	0.082
			400	0.196	0.164	0.126	0.100	0.074	0.058
0.8	0.4	3	225	0.462	0.390	0.610	0.414	0.596	0.496
			400	0.472	0.414	0.598	0.426	0.592	0.480
0.8	0.4	5	225	0.902	0.880	0.956	0.898	0.956	0.922
			400	0.914	0.868	0.960	0.906	0.958	0.930

Table 1: Rejection proportions of the null hypothesis for $\alpha = 0.05$. Non-scalar bandwidths.

				h				
σ	a_e	c	n	0.6	0.7	0.8	0.9	1
0.4	0.1	0	225	0.060	0.042	0.042	0.030	0.024
			400	0.044	0.040	0.030	0.022	0.016
0.4	0.1	3	225	0.454	0.408	0.394	0.392	0.398
			400	0.420	0.368	0.324	0.316	0.324
0.4	0.1	5	225	1.000	0.998	0.998	0.998	0.998
			400	1.000	0.998	0.998	0.996	0.994
0.4	0.2	0	225	0.086	0.058	0.048	0.032	0.024
			400	0.104	0.072	0.034	0.024	0.020
0.4	0.2	3	225	0.852	0.838	0.809	0.794	0.790
			400	0.886	0.862	0.834	0.822	0.816
0.4	0.2	5	225	1.000	1.000	1.000	1.000	1.000
			400	1.000	1.000	1.000	1.000	1.000
0.4	0.4	0	225	0.164	0.126	0.084	0.074	0.068
			400	0.172	0.130	0.096	0.080	0.064
0.4	0.4	3	225	0.978	0.976	0.974	0.970	0.970
			400	0.994	0.994	0.992	0.992	0.988
0.4	0.4	5	225	1.000	1.000	1.000	1.000	1.000
			400	1.000	1.000	1.000	1.000	1.000
0.6	0.1	0	225	0.060	0.042	0.042	0.034	0.024
			400	0.044	0.038	0.030	0.022	0.016
0.6	0.1	3	225	0.096	0.084	0.062	0.056	0.066
			400	0.072	0.046	0.036	0.030	0.030
0.6	0.1	5	225	0.640	0.606	0.574	0.558	0.568
			400	0.604	0.554	0.532	0.508	0.522
0.6	0.2	0	225	0.086	0.060	0.048	0.032	0.026
			400	0.104	0.066	0.032	0.022	0.016
0.6	0.2	3	225	0.418	0.362	0.314	0.282	0.272
			400	0.460	0.400	0.346	0.306	0.300
0.6	0.2	5	225	0.938	0.920	0.916	0.894	0.890
			400	0.944	0.942	0.926	0.916	0.916
0.6	0.4	0	225	0.158	0.126	0.084	0.074	0.068
			400	0.168	0.124	0.090	0.076	0.058
0.6	0.4	3	225	0.766	0.742	0.716	0.694	0.684
			400	0.810	0.776	0.740	0.716	0.708
0.6	0.4	5	225	1.000	0.998	0.998	0.998	0.998
			400	1.000	1.000	0.996	0.996	0.996
0.8	0.1	0	225	0.060	0.040	0.042	0.036	0.026
			400	0.044	0.038	0.030	0.022	0.016
0.8	0.1	3	225	0.062	0.028	0.016	0.012	0.018
			400	0.040	0.024	0.014	0.010	0.010
0.8	0.1	5	225	0.218	0.196	0.180	0.172	0.166
			400	0.164	0.128	0.102	0.110	0.124
0.8	0.2	0	225	0.086	0.060	0.048	0.034	0.026
			400	0.104	0.064	0.030	0.022	0.016
0.8	0.2	3	225	0.234	0.180	0.1386	0.118	0.116
			400	0.278	0.222	0.176	0.140	0.136
0.8	0.2	5	225	0.654	0.612	0.578	0.560	0.550
			400	0.698	0.644	0.586	0.568	0.560
0.8	0.4	0	225	0.158	0.124	0.084	0.074	0.068
			400	0.168	0.120	0.092	0.078	0.060
0.8	0.4	3	225	0.556	0.496	0.458	0.434	0.426
			400	0.572	0.542	0.494	0.474	0.460
0.8	0.4	5	225	0.928	0.920	0.906	0.888	0.874
			400	0.952	0.946	0.928	0.928	0.918

Table 2: Rejection proportions of the null hypothesis for $\alpha = 0.05$. Regression function (37).

σ	a_e	c	n	**H**					
				$\begin{pmatrix} 0.8 & 0 \\ 0 & 0.6 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 0.6 \end{pmatrix}$	$\begin{pmatrix} 0.6 & 0 \\ 0 & 0.8 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 0.8 \end{pmatrix}$	$\begin{pmatrix} 0.6 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0.8 & 0 \\ 0 & 1 \end{pmatrix}$
0.4	0.1	0	225	0.050	0.038	0.048	0.036	0.036	0.034
			400	0.038	0.032	0.034	0.026	0.030	0.024
0.4	0.1	3	225	0.328	0.272	0.520	0.354	0.538	0.422
			400	0.238	0.196	0.462	0.278	0.484	0.364
0.4	0.1	5	225	0.996	0.988	1.000	0.998	1.000	0.998
			400	0.992	0.986	1.000	0.990	1.000	0.998
0.4	0.2	0	225	0.062	0.042	0.064	0.032	0.054	0.038
			400	0.072	0.048	0.064	0.030	0.030	0.020
0.4	0.2	3	225	0.770	0.706	0.874	0.774	0.876	0.824
			400	0.798	0.728	0.904	0.796	0.912	0.848
0.4	0.2	5	225	1.000	1.000	1.000	1.000	1.000	1.000
			400	1.000	1.000	1.000	1.000	1.000	1.000
0.4	0.4	0	225	0.118	0.104	0.110	0.080	0.084	0.072
			400	0.136	0.114	0.136	0.084	0.110	0.072
0.4	0.4	3	225	0.970	0.952	0.982	0.966	0.984	0.972
			400	0.986	0.974	0.994	0.984	0.996	0.992
0.4	0.4	5	225	1.000	1.000	1.000	1.000	1.000	1.000
			400	1.000	1.000	1.000	1.000	1.000	1.000
0.6	0.1	0	225	0.048	0.042	0.046	0.036	0.042	0.034
			400	0.038	0.032	0.034	0.026	0.030	0.022
0.6	0.1	3	225	0.052	0.044	0.110	0.054	0.124	0.072
			400	0.026	0.014	0.096	0.026	0.108	0.040
0.6	0.1	5	225	0.496	0.442	0.684	0.530	0.708	0.612
			400	0.452	0.378	0.680	0.470	0.700	0.562
0.6	0.2	0	225	0.062	0.042	0.066	0.032	0.054	0.038
			400	0.070	0.046	0.064	0.028	0.030	0.016
0.6	0.2	3	225	0.278	0.214	0.456	0.258	0.466	0.332
			400	0.308	0.240	0.486	0.278	0.486	0.356
0.6	0.2	5	225	0.878	0.846	0.946	0.892	0.948	0.916
			400	0.900	0.856	0.954	0.900	0.954	0.934
0.6	0.4	0	225	0.116	0.102	0.108	0.080	0.084	0.072
			400	0.134	0.110	0.132	0.078	0.106	0.070
0.6	0.4	3	225	0.664	0.606	0.806	0.654	0.798	0.732
			400	0.706	0.646	0.828	0.672	0.826	0.750
0.6	0.4	5	225	0.994	0.988	0.998	0.994	0.998	0.998
			400	0.996	0.996	1.000	0.996	1.000	1.000
0.8	0.1	0	225	0.050	0.042	0.048	0.036	0.042	0.036
			400	0.038	0.032	0.034	0.026	0.030	0.022
0.8	0.1	3	225	0.010	0.010	0.078	0.012	0.080	0.018
			400	0.008	0.006	0.050	0.008	0.054	0.018
0.8	0.1	5	225	0.134	0.118	0.262	0.158	0.274	0.200
			400	0.066	0.054	0.222	0.086	0.232	0.138
0.8	0.2	0	225	0.062	0.040	0.064	0.032	0.054	0.040
			400	0.068	0.042	0.064	0.028	0.030	0.016
0.8	0.2	3	225	0.132	0.096	0.262	0.108	0.252	0.148
			400	0.168	0.110	0.302	0.124	0.282	0.182
0.8	0.2	5	225	0.520	0.456	0.684	0.518	0.684	0.610
			400	0.546	0.450	0.720	0.518	0.728	0.612
0.8	0.4	0	225	0.114	0.102	0.110	0.080	0.084	0.074
			400	0.134	0.112	0.130	0.080	0.106	0.072
0.8	0.4	3	225	0.432	0.354	0.596	0.406	0.570	0.464
			400	0.492	0.418	0.590	0.440	0.588	0.516
0.8	0.4	5	225	0.884	0.834	0.934	0.870	0.932	0.908
			400	0.918	0.872	0.958	0.904	0.956	0.930

Table 3: Rejection proportions of the null hypothesis for $\alpha = 0.05$. Regression function (37).

2.4 Random design

The methodology is now illustrated with covariate variables generated from a random design. As in the main paper, the regression function $m(X_1, X_2) = 2 + X_1 + X_2 + cX_1^3$ is considered. In this case, for each value of c (being c equal to 0 or 5), 500 samples of sizes $n = 225$ and 400 are uniformly sampled in the unit square. The random errors ε_i are normally distributed with zero mean and isotropic exponential covariance function (36), with $\sigma = 0.4$, 0.8, and $a_e = 0.1$, 0.4. No nugget effect is considered. Table 4 shows the rejection proportions of the null hypothesis, for $\alpha = 0.05$, considering that the bandwidth is taken as a diagonal matrix $\mathbf{H} = \text{diag}(h, h)$, and different values of h are chosen, $h = 0.6, 0.7, 0.8, 0.9, 1$. Similar conclusions as in the case of considering a fixed design can be deduced.

				h				
σ	a_e	c	n	0.6	0.7	0.8	0.9	1
0.4	0.1	0	225	0.066	0.056	0.036	0.028	0.022
			400	0.080	0.068	0.058	0.048	0.042
0.4	0.1	5	225	1.000	1.000	1.000	1.000	1.000
			400	1.000	1.000	1.000	1.000	1.000
0.4	0.4	0	225	0.144	0.100	0.082	0.060	0.052
			400	0.146	0.118	0.086	0.068	0.056
0.4	0.4	5	225	1.000	1.000	1.000	1.000	1.000
			400	1.000	1.000	1.000	1.000	1.000
0.8	0.1	0	225	0.072	0.056	0.036	0.030	0.024
			400	0.080	0.068	0.058	0.048	0.044
0.8	0.1	5	225	0.916	0.890	0.870	0.860	0.858
			400	0.954	0.946	0.944	0.944	0.948
0.8	0.4	0	225	0.142	0.112	0.100	0.090	0.076
			400	0.160	0.122	0.086	0.062	0.054
0.8	0.4	5	225	1.000	1.000	1.000	1.000	1.000
			400	1.000	1.000	1.000	1.000	1.000

Table 4: Rejection proportions of the null hypothesis for $\alpha = 0.05$. Random design.

2.5 Nugget effect

Finally, a nugget effect is included in the dependence model. Recall that in the previous frameworks the nugget effect was zero. In this case, the model considered is similar to the one of the main paper: the regression function is the same, $m(X_1, X_2) = 2 + X_1 + X_2 + cX_1^3$ (the data are generated on a bidimensional regular grid in the unit square, and c is considered equal to 0 or 5). However, a nugget effect is included in the dependence structure. Then, the random errors ε_i are normally distributed with zero mean and isotropic exponential covariance function: $\text{Cov}(\varepsilon_i, \varepsilon_j) = c_e\{\exp(-\|\mathbf{X}_i - \mathbf{X}_j\|/a_e)\}$, if $\|\mathbf{X}_i - \mathbf{X}_j\| \neq 0$, where $c_e = \sigma^2 - c_0$ is the partial sill, with $\sigma = 0.4$ and nugget effect c_0 being 20% and 50% of the total variance σ^2. Two values for the practical range are considered, $a_e = 0.1$ and 0.4. Table 5 shows the rejection proportions of the null hypothesis, for $\alpha = 0.05$, considering that the bandwidth is taken as a diagonal matrix $\mathbf{H} = \text{diag}(h, h)$, and different values of h are chosen, $h = 0.6, 0.7, 0.8, 0.9, 1$. It can be observed that the performance of the test is satisfactory, with similar results to those in the previous scenarios. As the nugget is larger, the bandwidth value should be smaller.

c_0	σ	a_e	c	n	h 0.6	0.7	0.8	0.9	1
	0.4	0.1	0	225	0.078	0.060	0.042	0.030	0.026
				400	0.052	0.038	0.028	0.016	0.010
20%	0.4	0.1	5	225	1.000	1.000	1.000	1.000	1.000
				400	1.000	1.000	1.000	1.000	1.000
	0.4	0.1	0	225	0.074	0.056	0.030	0.020	0.020
				400	0.028	0.016	0.014	0.012	0.012
50%	0.4	0.1	5	225	1.000	1.000	1.000	1.000	1.000
				400	1.000	1.000	1.000	1.000	1.000
	0.4	0.4	0	225	0.052	0.048	0.036	0.036	0.030
				400	0.044	0.040	0.032	0.026	0.020
20%	0.4	0.4	5	225	1.000	1.000	1.000	1.000	1.000
				400	1.000	1.000	1.000	1.000	1.000
	0.4	0.4	0	225	0.062	0.050	0.044	0.038	0.026
				400	0.024	0.024	0.020	0.020	0.014
50%	0.4	0.4	5	225	1.000	1.000	1.000	1.000	1.000
				400	1.000	1.000	1.000	1.000	1.000

Table 5: Rejection proportions of the null hypothesis for $\alpha = 0.05$. Nugget effect.

References

Härdle, W. and E. Mammen (1993). Comparing nonparametric versus parametric regression fits. *Ann. Stat. 21*, 1926–1947.

Kim, T. Y., J. Ha, S. Y. Hwang, C. Park, and Z.-M. Luo (2013). Central limit theorems for reduced U-statistics under dependence and their usefulness. *Aust. N. Z. J. Stat. 55*, 387–399.

Liu, X. H. (2001). *Kernel smoothing for spatially correlated data*. Ph. D. thesis, Department of Statistics, Iowa State University.