Published in final edited form as:

J Stat Plan Inference. 2010 October 1; 140(10): 2875–2884. doi:10.1016/j.jspi.2010.03.011.

On nonparametric comparison of images and regression surfaces

Xiao-Feng Wang^{a,b,*} and Deping Ye^c

Xiao-Feng Wang: wangx6@ccf.org; Deping Ye: deping.ye@case.edu

^a Department of Quantitative Health Sciences/Biostatistics, Cleveland Clinic Foundation, Cleveland, OH 44195, USA

^b Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH 44106, USA

^c Department of Mathematics, Case Western Reserve University, Cleveland, OH 44106, USA

Abstract

Multivariate local regression is an important tool for image processing and analysis. In many practical biomedical problems, one is often interested in comparing a group of images or regression surfaces. In this paper, we extend the existing method of testing the equality of nonparametric curves by Dette and Neumeyer (2001) and consider a test statistic by means of an $\mathcal{L}^2$ -distance in the multi-dimensional case under a completely heteroscedastic nonparametric model. The test statistic is also extended to be used in the case of spatial correlated errors. Two bootstrap procedures are described in order to approximate the critical values of the test depending on the nature of random errors. The resulting algorithms and analyses are illustrated from both simulation studies and a real medical example.

Keywords

Multivariate local regression; images; equality; regression surfaces; wild bootstrap; heteroscedastic model; spatial correlated errors

1. Introduction

Medical images are increasingly used in health care and biomedical research and a wide range of imaging modalities are now available. Statistical image analysis, hence, becomes an active research area. Nonparametric regression techniques have been broadly applied to image analysis, including image reconstruction, denoising and interpolation. An image can be considered as a surface of the image intensity at each pixel. A regression surface from a noisy image is often fitted by local smoothing procedures (Qiu, 1998; Takeda et al., 2007). Based on the nonparametric modeling framework, one object of primary interest is to compare a set of smoothed images or regression surfaces. For example, one often exams the equality of two or more images under different clinical conditions in medical applications.

*Corresponding author: Xiao-Feng Wang, Ph.D., Department of Quantitative Health Sciences/Biostatistics, Cleveland Clinic Foundation, 9500 Euclid Avenue/JJN3, Cleveland, OH 44195.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Nonparametric comparison of a set of regression curves has been paid considerable attentions in both theoretical and applied regression analysis. Much effort has been devoted to this problem in the literature. Hall and Hart (1990); King et al. (1991) had early considerations of the problem, where they discussed a completely nonparametric homoscedastic model in the case of equal design points. While Kulasekera (1995) proposed several alternative tests in the case of unequal design points, Young and Bowman (1995) generalized the one-way analysis of variance (ANOVA) to the nonparametric regression setting. Hardle and Mammen (1990) considered a method based on a weighted $\mathcal{L}^2$ -distance for semiparametric comparison of regression curves. Dette and Neumeyer (2001) discussed three methods using nonparametric estimators of the regression function in the problem of testing the equality of k regression curves from independent samples. Their first test was based on a linear combination of variance estimators; The second approach was an ANOVA-type method; and the third test compared the differences between the estimates of the individual regression curves by means of an $\mathcal{L}^2$ -distance. Recent contributions on this problem include Pardo-Fernandez et al. (2007); Pardo-Fernandez (2007) among others. Several new testing statistics have been discussed such as Kolmogorov-Smirnov type and Cramer-von Mises type statistics. A good and recent review on this topic can be found in Neumeyer and Dette (2003).

Nonparametric comparison of different images or regression surfaces relates to nonparametric analysis of covariance with multiple covariates. It is important to extend the methods for nonparametric curve comparison to the multi-dimensional case with applications to image analysis. Recently, Bowman (2006) suggested a generalization of the ANOVA-type test by Young and Bowman (1995) to compare regression surfaces. Under the assumptions of equal homoscedastic variances in all groups and normal distributed errors, Bowman (2006) proposed a χ^2 -approximation of the corresponding test statistic under the null hypothesis. In this paper, we discuss the comparison of regression surfaces under a general model which does not require any additional assumptions (such as homoscedasticity or normality of errors, equal design points). In Section 2, we review the classic framework of local regression for image data and suggest a generalization of Dette and Neumeyer (2001)'s test, a test procedure based on an $\mathcal{L}^2$ -distance of regression surfaces under a general heteroscedastic model. The asymptotic results of the proposed test statistic are presented here. We also extend the test statistic to be used in the case of spatial correlated errors. Two bootstrap procedures are described in order to approximate the critical values of the test. In Section 3, we present numerical examples. Simulation studies are conducted to investigate the finite sample properties of the proposed test. A real data analysis is performed to illustrate the use of our method. The paper ends with the concluding remarks in Section 4.

2. Testing the the equality of images and regression surfaces

2.1. Local surface model for images

An image can be represented as a function $m(\cdot)$ from a plane $\mathbb{R}^2$ to p -variate space $\mathbb{R}^p$, where the value of $m(\cdot)$ in the j th coordinate represents its intensity. The dimension of $m(\cdot)$, p , can be greater than 1. For instance, $p = 3$ for a natural color image, representing the three primary colors. For simplicity and clarity we will treat only the case $p = 1$, while the methodology is similar in other cases. In image application, the function $m(\cdot)$ is assumed to be a smooth function of $\mathbf{X} \in \mathbb{R}^2$. We only have discrete data on the model with the $\mathbf{X}$ variable restricted to a regular or irregular grid and the intensity values observed are often measured with noise. In general, this model can be formalized as

$$Y_j = m(\mathbf{X}_j) + \sigma_i(\mathbf{X}_j)\varepsilon_j, \quad j=1, \dots, n. \quad (2.1)$$

where ε_j are independent and identically distributed random variables, which represents random errors in the observations. We further assume that ε_j have zero mean and finite variance 1.

Using the data $(\mathbf{X}_j, Y_j), j = 1, \dots, n$, we want to construct a “denoised” image, an estimator of the regression function $m(\cdot)$ which is the conditional expectation of the dependent variable Y given the independent variable $\mathbf{X}$,

$$m(\mathbf{x}) = E(Y|\mathbf{X}=\mathbf{x}).$$

Local smoothing method is an important tool in image processing and analysis (Wand and Jones, 1995; Takeda et al., 2007). We now describe how to construct a smoothed image by local approximations of $m(\cdot)$, using a least-squares method with kernel weights.

The Taylor's expansion of the regression function at $\mathbf{X}_j$ implies that,

$$m(\mathbf{X}_j) \approx \beta_0 + \beta_1^T (\mathbf{X}_j - \mathbf{x}) + \beta_2^T \text{vech}\{(\mathbf{X}_j - \mathbf{x})^T (\mathbf{X}_j - \mathbf{x})\} + \dots,$$

where $\text{vech}(\cdot)$ returns the vector obtained by eliminating all supradiagonal elements of the square matrix and stacking the result one column above the other. $\beta_0 = m(\mathbf{x})$ is the pixel value of interest and the vectors β_1 and β_2 are

$$\beta_1 = \left[\frac{\partial m(\mathbf{x})}{\partial x_1}, \frac{\partial m(\mathbf{x})}{\partial x_2} \right]^T, \beta_2 = \frac{1}{2} \left[\frac{\partial^2 m(\mathbf{x})}{\partial x_1^2}, 2 \frac{\partial^2 m(\mathbf{x})}{\partial x_1 \partial x_2}, \frac{\partial^2 m(\mathbf{x})}{\partial x_2^2} \right]^T.$$

Estimating $\beta = [\beta_0, \beta_1^T, \beta_2^T, \dots]^T$ can be through solving the following least squares problem,

$$\min_{\beta} \sum_{j=1}^n \{Y_j - \beta_0 - \beta_1^T (\mathbf{X}_j - \mathbf{x}) - \beta_2^T \text{vech}\{(\mathbf{X}_j - \mathbf{x})^T (\mathbf{X}_j - \mathbf{x})\} - \dots\}^2 K_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_j).$$

The weight function (kernel function) $K_{\mathbf{H}}(\mathbf{x}) = \det(\mathbf{H})^{-1} K(\mathbf{H}^{-1}\mathbf{x})$ is defined on the multivariate space, hence observations close to a fitting point $\mathbf{x}$ receive large weights. $\mathbf{H}$ is a bandwidth matrix which is symmetric positive-definite and $\det(\mathbf{H})$ is the determinant of the matrix $\mathbf{H}$. The local least squares estimator of $m(\mathbf{x})$ is

$$\widehat{m}_{\mathbf{H}}(\mathbf{x}) = \mathbf{e}_1^T (\tilde{\mathbf{X}}^T \mathbf{W} \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \mathbf{W} \mathbf{Y}, \quad (2.2)$$

where $\mathbf{e}_1$ is a column vector with the first element equal to one and the rest equal to zero; $\mathbf{Y} = [Y_1, \dots, Y_n]^T$, $\mathbf{W} = \text{diag}\{K_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_1), \dots, K_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_n)\}$, and

$$\tilde{\mathbf{X}} = \begin{bmatrix} 1 & (\mathbf{X}_1 - \mathbf{x})^T & \text{vech}^T\{(\mathbf{X}_1 - \mathbf{x})(\mathbf{X}_1 - \mathbf{x})^T\} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (\mathbf{X}_n - \mathbf{x})^T & \text{vech}^T\{(\mathbf{X}_n - \mathbf{x})(\mathbf{X}_n - \mathbf{x})^T\} & \cdots \end{bmatrix}.$$

When we consider a local constant approximation of $m(\mathbf{X}_j)$, (2.2) is indeed the *multivariate Nadaraya-Watson estimator*,

$$\hat{m}_{\mathbf{H}}(\mathbf{x}) = \frac{\sum_{j=1}^n K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x}) Y_j}{\sum_{j=1}^n K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x})}. \quad (2.3)$$

There are two typical approaches for constructing multivariate kernel $K(\cdot)$. One is the applicable preferred product kernels $K(\mathbf{x}) = \prod_{s=1}^2 k_s(x_s)$ where each $k_s(\cdot)$ denotes the univariate kernel function. Another is the spherically or radially symmetric kernel $K(\mathbf{x}) = rK(\|\mathbf{x}\|)$ where $r = (\int K(\|\mathbf{x}\|) d\mathbf{x})^{-1}$ is the normalization constant and $\|\mathbf{x}\|^2 = \sum_{s=1}^2 x_s^2$. The easiest bandwidth matrix is $\mathbf{H} = h\mathbf{I}$ where $\mathbf{I}$ denotes the 2×2 identity matrix, assuming that the independent variables have the same scale. A more complex bandwidth matrix can be $\mathbf{H} = h\Sigma$, where Σ is a positive definite matrix independent of h , and $\det(\mathbf{H}) = h^2 \det(\Sigma)$. For a complete discussion about multivariate local regression, we refer to the monographs by Wand and Jones (1995); Hardle et al. (2004).

2.2. Comparing nonparametric surfaces

In many medical applications, such as the NMES study (see the real data application in section 3), a question of particular interest is to examine the equality of two or more images under different clinical conditions. This problem is related to the comparison of nonparametric regression surfaces across several groups. It is noted that the variances of random errors of images across different conditions can be different. A suitable model is the following general heteroscedastic model that can be written as,

$$Y_{ij} = m_i(\mathbf{X}_{ij}) + \sigma_i(\mathbf{X}_{ij})\varepsilon_{ij}, \quad i=1, \dots, L; \quad j=1, \dots, n_i. \quad (2.4)$$

Here we assume that ε_{ij} are independent identically distributed errors with mean 0, variance 1 and finite 4-th moments. $m_i(\cdot): \mathbb{R}^2 \rightarrow \mathbb{R}$ are unknown but smooth regression functions, and $\sigma_i(\cdot): \mathbb{R}^2 \rightarrow \mathbb{R}_+$ are the smooth variance functions. Without loss of generality, we assume $\mathbf{X} \in S = [0, 1] \times [0, 1]$. We are interested in the problem of testing the equality of the regression surfaces, that is,

$$\begin{aligned} H_0: m_i(\mathbf{x}) &\equiv m_k(\mathbf{x}), \quad \text{for all } i, k=1, \dots, L \quad \text{v.s.} \\ H_1: m_i(\mathbf{x}) &\neq m_k(\mathbf{x}), \quad \text{for some } i \neq k. \end{aligned} \quad (2.5)$$

We use N to denote $N = \sum_{i=1}^L n_i$ and assume there exist $0 < \kappa_i < 1$, $i = 1, \dots, L$ such that

$$\frac{n_i}{N} = \kappa_i + O(1/N). \quad (2.6)$$

$O(1/N)$ means that $\lim_{N \rightarrow \infty} N |O(1/N)| = C$ for some constant C with $0 < C < \infty$. Following Hardle and Mammen (1993) and Dette and Neumeyer (2001), an obvious test of the hypothesis (2.5) could be obtained from a comparison of the differences between the estimates of the individual regression surfaces by means of an $\mathcal{L}^2$ -distance. To this end, we consider the statistic

$$T_N = \sum_{1 \leq i < k \leq L} \int [\hat{m}_i(\mathbf{x}) - \hat{m}_k(\mathbf{x})]^2 \omega(\mathbf{x}) d\mathbf{x}, \quad (2.7)$$

where $\omega(\cdot)$ is a (smooth) positive weight function, and $\hat{m}_i(x)$ and $\hat{m}_k(x)$ ($1 \leq i < k \leq L$) denote the local smooth estimators for the i -th and the k -th group data computed from (2.2). The weight function is often chosen by the experienced researcher and serves to trim the boundaries or regions of sparse data. For example, the simplest case is $\omega(x) = 1$. If $\omega(\mathbf{x})$ is equal to $f_{\mathbf{X}}(\mathbf{x})$, the density of $\mathbf{X}$, and $n_1 = \dots = n_L = n$, one could take the empirical version of (2.7)

$$\tilde{T}_N = \frac{1}{n} \sum_{1 \leq i < k \leq L} \sum_{j=1}^n (\hat{m}_i(\mathbf{X}_j) - \hat{m}_k(\mathbf{X}_j))^2$$

as the test statistic.

T_N in (2.7) is the multi-dimensional generalization for the test statistic $T_N^{(3)}$ in Dette and Neumeyer (2001). We propose to use it in image applications rather than use the generalization of $T_N^{(1)}$ and $T_N^{(2)}$ in Dette and Neumeyer (2001), because it is more meaningful and easier to be constructed in image applications. Moreover, T_N can apply to the case when images are blurred with spatial correlated noises, while the two other statistics seem not be able to work directly with such a case.

For the sake of simplicity and clarity, we present the asymptotic results of the statistic T_N for $L = 2$. The asymptotic results for $L > 2$ can be obtained similarly. Here we choose the same bandwidth matrix $\mathbf{H} = h\Sigma$ and the same kernel function $K(\cdot)$ for different group data. As before, we let $K_{\mathbf{H}}(\mathbf{x}) = \det(\mathbf{H})^{-1} K(\mathbf{H}^{-1}\mathbf{x})$. We further assume that the bandwidth matrix $\mathbf{H}$ satisfies

$$N \det(\mathbf{H}) \rightarrow \infty \text{ and } N \det^2(\mathbf{H}) \rightarrow 0, \quad (2.8)$$

which is equivalent to $Nh^2 \rightarrow \infty$ and $Nh^4 \rightarrow 0$. The asymptotic properties of the statistic T_N are listed in the following theorem. Since observed data are often on a regular grid with a fixed design in image applications, the theorem is based on a fixed design. It is noted that the part (ii) of the theorem will be different if we consider a random design; see Dette (1999) for the difference between these two.

Theorem 2.1—Suppose that conditions 2.6 and 2.8 are satisfied and $K(\cdot)$ is a second-order kernel.

- i. If $m_1(\mathbf{x}) \equiv m_2(\mathbf{x})$ holds, then the statistic T_N satisfies that

$$N \sqrt{\det(\mathbf{H})} \left(T_N - \frac{\alpha}{N \det(\mathbf{H})} \right) \rightarrow \mathcal{N}(0, \beta^2) \text{ in distribution as } N \sqrt{\det(\mathbf{H})} \rightarrow \infty \text{ where}$$

$$\alpha = \|K\|_2^2 \int \left[\frac{\sigma_1^2(\mathbf{x})}{\kappa_1 f_1(\mathbf{x})} + \frac{\sigma_2^2(\mathbf{x})}{\kappa_2 f_2(\mathbf{x})} \right] \omega(\mathbf{x}) d\mathbf{x},$$

$$\beta^2 = 2 \|K * K\|_2^2 \int \left(\frac{\sigma_1^2(\mathbf{x})}{\kappa_1 f_1(\mathbf{x})} + \frac{\sigma_2^2(\mathbf{x})}{\kappa_2 f_2(\mathbf{x})} \right)^2 \omega^2(\mathbf{x}) d\mathbf{x}.$$

Here $f_i(\cdot)$, $i = 1, 2$ are the density functions of $\mathbf{X}_{ij}$, $\|K\|_2^2 = \int K(\mathbf{x})^2 d\mathbf{x}$ and $(K * K)(\mathbf{x}) = \int K(\mathbf{x} - \mathbf{y})K(\mathbf{y}) d\mathbf{y}$.

- ii. Under the alternative $m_1(\mathbf{x}) \neq m_2(\mathbf{x})$, the statistic T_N satisfies that

$$\sqrt{N}(T_N - M^2) \rightarrow \mathcal{N}(0, \gamma^2) \text{ in distribution as } N \rightarrow \infty$$

where

$$M^2 = \int [m_2(\mathbf{x}) - m_1(\mathbf{x})]^2 \omega(\mathbf{x}) d\mathbf{x},$$

$$\gamma^2 = 4 \int \left\{ [m_1(\mathbf{x}) - m_2(\mathbf{x})]^2 \omega^2(\mathbf{x}) \left(\frac{\sigma_1^2(\mathbf{x})}{\kappa_1 f_1(\mathbf{x})} + \frac{\sigma_2^2(\mathbf{x})}{\kappa_2 f_2(\mathbf{x})} \right) \right\} d\mathbf{x}.$$

The key rationale of the proof for Theorem 2.1 remains the same as in Dette and Neumeyer (2001). The proof heavily relies on the Taylor extension formula, central limit theorem, law of large number, Holder inequalities, Fubini's theorem, and Chebyshev inequality. In particular, for part (i), the asymptotic normality in the case of $m_1 = m_2$, one needs to rewrite T_N as a generalized quadratic-form and proves the asymptotic normality by using the theorem 2.1 in De Jong (1987). The part (ii) is obtained by checking the Lyapunov condition. The detailed proof can be found in our technical report at stat.case.edu/~xfwang/paper/TestSurfaces_Techrep.pdf.

2.3. Bootstrap approximation

In practice, the asymptotic distribution of the test statistic T_N is difficult to be used for the test because of the slow convergence of T_N towards the normal distribution. The analogous problems occur in nonparametric comparison of regression curves (Neumeyer and Dette, 2003; Pardo-Fernandez et al., 2007). However, theorem 2.1 reassures us to construct the reference distribution by bootstrapping. Similar to Hardle and Mammen (1993) and Dette and Neumeyer (2001), we propose the popular resampling method so-called *wild bootstrap* procedure introduced by Wu (1986) to our test (2.7). An advantage of the procedure is that it allows for a heterogeneous variance in the residuals. The algorithm can be described with the following steps.

Algorithm 2.1—Wild bootstrap for nonparametric surface comparison with heterogeneous errors.

1. Estimate $\hat{m}_i(\mathbf{x})$ and compute the test statistic T_N based on (2.7); then estimate the common regression surface $\hat{m}(\mathbf{x})$ from the total sample $\mathbf{X}_{ij}$ and construct the residuals $\hat{\epsilon}_{ij} = Y_{ij} - \hat{m}(\mathbf{X}_{ij})$.
2. For each $\mathbf{X}_{ij}$, draw a bootstrap residual $\hat{\epsilon}_{ij}^*$ such that $E(\hat{\epsilon}_{ij}^*) = 0$, $E(\hat{\epsilon}_{ij}^{*2}) = \hat{\epsilon}_{ij}^2$ and $E(\hat{\epsilon}_{ij}^{*3}) = \hat{\epsilon}_{ij}^3$. A typical choice is drawing $\hat{\epsilon}_{ij}^*$ from the two-point distribution with

probability mass at $a = \frac{1-\sqrt{5}}{2}\widehat{\varepsilon}_{ij}$, $b = \frac{1+\sqrt{5}}{2}\widehat{\varepsilon}_{ij}$, occurring with probabilities $q = (5 + \sqrt{5})/10$ and $1 - q$, respectively.

3. Generate a bootstrap sample $\{(\mathbf{X}_{ij}, Y_{ij}^*)\}$ by setting

$$Y_{ij}^* = \widehat{m}(\mathbf{X}_{ij}) + \widehat{\varepsilon}_{ij}^*.$$

4. From this sample, compute the bootstrap regression surfaces $\widehat{m}_i^*(\mathbf{x})$ and the test statistic T_N^* in the same way as the original T_N is computed.
5. Repeat steps (2) to (4) B times and use the B generated test statistics T_N^* to determine the quantiles of the test statistic under the null hypothesis. For the test at level α , the null hypothesis is rejected if T_N is greater than the corresponding quantile of the bootstrap distribution of T_N .

Remark 2.1—As we will show in the next section, the selection of the bandwidths in the testing procedure does not have a big impact on the rejection probabilities. In practice, we propose to select the bandwidth using the simple extension of bias-corrected Akaike Information Criterion (AIC_C) criterion (Hurvich et al., 1998; Bowman, 2006). Hurvich et al. (1998) show that the use of AIC_C avoids the large variability and tendency to undersmooth when generalized cross-validation (GCV) or the Akaike Information Criterion (AIC) are used to choose the smoothing parameter. The AIC_C was originally used for the nonparametric curve estimation, but it is easy to be extended for surface estimation. One can use a product kernel for regression surface estimation,

$$K_H(\mathbf{X}_i - \mathbf{x}) = \frac{1}{h_1 h_2} k\left(\frac{X_{1i} - x_1}{h_1}\right) k\left(\frac{X_{2i} - x_2}{h_2}\right), \quad (2.9)$$

where k denotes a univariate kernel function. An appropriate smoothing parameters is to set $(h_1, h_2) = (h_0 \hat{\tau}_1, h_0 \hat{\tau}_2)$ and $\hat{\tau}_2$ are estimated standard deviations of the two predictors. The optimal $\hat{h}_0$ is chosen to minimize

$$AIC_C(h_0) = \log(\widehat{\sigma}^2) + 1 + \frac{2(\nu + 1)}{n - \nu - 2},$$

where ν is the trace of the hat matrix of the model and $\widehat{\sigma}^2 = \sum_{i=1}^n (y_i - \widehat{m}_H(\mathbf{X}_i))^2 / n$.

In many image applications, it often occurs that the random errors of an image are spatially correlated (Cressie, 1993). Our proposed test statistic is also applicable in the case of spatial correlated errors. In such a case, the error term $\sigma_i(\mathbf{X}_{ij})\varepsilon_{ij}$ in the model (2.4) becomes η_{ij} , where $E(\eta_{ij}|\mathbf{X}_{ij}) = 0$, $Var(\eta_{ij}|\mathbf{X}_{ij}) = \sigma_i^2$ and $Cov(\eta_{ij}, \eta_{ik}|\mathbf{X}_{ij}, \mathbf{X}_{ik}) = \sigma_i^2 \rho_i(\mathbf{X}_{ij} - \mathbf{X}_{ik})$ with $\rho_i(\mathbf{x})$ continuous, satisfying $\rho_i(\mathbf{0}) = 1$, $\rho_i(\mathbf{x}) = \rho_i(-\mathbf{x})$, and $\rho_i(\mathbf{x}) \leq 1$ for all $\mathbf{x}$. The wild bootstrap in algorithm 2.1 is only designed for the case of heterogeneous errors. We suggest the following “indirect” method when image data subject with spatial correlated errors.

Algorithm 2.2—Indirect bootstrap for nonparametric surface comparison with spatial correlated errors.

1. Construct the preliminary local estimates $\hat{m}_i(\cdot)$ and $\hat{m}(\cdot)$ with spatially correlated errors following Francisco-Fernandez and Opsomer (2005).
2. For $i = 1, \dots, L$, use the i th residuals $\hat{\eta}_i = (\hat{\eta}_{i1}, \dots, \hat{\eta}_{in_i})^T$ to model the semivariogram of the residuals and then yields the estimate of the corresponding covariance matrix $\hat{\mathbf{R}}_i$ for each sample.
3. Calculate the Cholesky decomposition $\hat{\mathbf{L}}_i$ such that $\hat{\mathbf{R}}_i = \hat{\mathbf{L}}_i \hat{\mathbf{L}}_i^T$ and obtain $\hat{\mathbf{e}}_i = (\hat{e}_1, \dots, \hat{e}_{n_i})^T = \hat{\mathbf{L}}_i^{-1} \hat{\eta}_i$; then calculate the “whitened” residuals $\tilde{\mathbf{e}}_i$ by computing $\tilde{\mathbf{e}}_i = \hat{\mathbf{e}}_i - \frac{1}{n} \sum_{j=1}^{n_i} \hat{e}_{ij}$.
4. Draw the bootstrap “whitened” residuals $\tilde{\mathbf{e}}_i^*$ and generate a bootstrap sample by introducing the correlation structure to the bootstrap “whitened” residuals, $\tilde{\eta}_i^* = \hat{\mathbf{L}}_i \tilde{\mathbf{e}}_i^*$; and then define new responses under the null hypothesis $Y_{ij}^* = \hat{m}(\mathbf{X}_{ij}) + \tilde{\eta}_{ij}^*$.
5. The distribution of the test statistic T_N is estimated by the empirical distribution of the statistic in B bootstrap samples.

3. Numerical examples

3.1. Simulation studies

In this subsection, we investigate the practical behavior of the proposed bootstrap procedure by simulations. We considered the following functions:

- a. $m_1(\mathbf{x}) = m_2(\mathbf{x}) = x_1 + x_2$
- b. $m_1(\mathbf{x}) = m_2(\mathbf{x}) = \sin(2\pi x_1) + \cos(2\pi x_1)$
- c. $m_1(\mathbf{x}) = m_2(\mathbf{x}) + 2x_1^2 + 3x_2^2$
- d. $m_1(\mathbf{x}) = m_2(\mathbf{x}) = e^{-x_1^2 - x_2^2}$
- e. $m_1(\mathbf{x}) = x_1 + x_2$ $m_2(\mathbf{x}) = -x_1 - x_2$
- f. $m_1(\mathbf{x}) = \sin(2\pi x_1) + \cos(2\pi x_2)$ $m_2(\mathbf{x}) = \sin(2\pi x_1) + \cos(2\pi x_2) + x_1$
- g. $m_1(\mathbf{x}) = 2x_1^2 + 3x_2^2$ $m_2(\mathbf{x}) = 2x_1^2 + 3x_2^2 + \sin(2\pi x_1)$
- h. $m_1(\mathbf{x}) = e^{-x_1^2 - x_2^2}$ $m_2(\mathbf{x}) = e^{-x_1^2 - x_2^2} + \sin(2\pi x_1)$

The cases (a) ~ (d) correspond to the null hypothesis of equal regression surfaces, while the cases (e) ~ (h) correspond to the alternative hypothesis. In all cases, the covariates x_1 and x_2 were generated from a regular grid of values in $S = [0, 1] \times [0, 1]$ with a fixed design. In each case, we considered both homoscedastic errors and heteroscedastic errors. We set $\omega(x) = 1$ in the statistic T_N . The variance functions in the case of homoscedastic errors were $\sigma_1^2(\mathbf{x}) = 0.5$ and $\sigma_2^2(\mathbf{x}) = 0.25$; the variance functions in the case of heteroscedastic errors were

$$\sigma_1^2(\mathbf{x}) = \sigma_2^2(\mathbf{x}) = \frac{1}{3} \exp\left(\sqrt{x_1^2 + x_2^2}\right).$$

The distributions of ε_1 and ε_2 were the standard normal distribution.

Simulation requires to specify the multidimensional kernel function. We used the product kernel (2.9) with a common bandwidth h , where its uni-variate kernel was the Epanechnikov kernel $k(x) = 3(1 - x^2)/4 \cdot I(|x| \leq 1)$. Following Dette and Neumeyer (2001), we considered simple rule-of-thumb (ROT) bandwidth selectors in our simulation. For each individual surface,

$$\widehat{h}_i = C_0 \left(\frac{\int_s \sigma_i^2(\mathbf{x}) d\mathbf{x}}{n_i} \right)^{3/10}, \quad i=1, 2;$$

for the common surface from the total sample in Algorithm 2.1,

$$\widehat{h} = C_0 \left(\frac{n_1 \int_s \sigma_1^2(\mathbf{x}) d\mathbf{x} + n_2 \int_s \sigma_2^2(\mathbf{x}) d\mathbf{x}}{(n_1 + n_2)^2} \right)^{3/10},$$

where C_0 is a certain constant. We performed 1000 trials with the bootstrap size $B = 250$ for every case of the simulation studies.

Our first two studies investigate the approximations of the level and the power by the wild bootstrap version of the test in Algorithm 2.1. We compared the proposed method (denoted by T_N) with the ANOVA-type test (denoted by T_{ANOVA}) using a χ^2 -approximation by Bowman (2006) for both homoscedastic errors and heteroscedastic errors. C_0 in the ROT bandwidth selectors was set to 1. Table 1 and 2 display the proportion of rejections in 1000 trials for sample sizes $(n_1, n_2) = (49, 49), (100, 49), (225, 225)$. The significant levels are $\alpha = 5\%$ and $\alpha = 10\%$. We observe that the level based on T_N is well approximated in most cases and the behavior of the power based on T_N is very good. In Table 1, the tests based on T_{ANOVA} seem to be conservative, especially as the sample size is small. In Table 2, the power obtained in both tests is better for larger sample sizes. The results listed also demonstrate a poor performance of T_{ANOVA} under heteroscedasticity. T_N receives a substantial improvement with respect to the power comparing with T_{ANOVA} , when errors are heteroscedastic.

We then studied the effects of the selection of the bandwidth on the proposed test. This is illustrated in Table 3, which shows results under the null hypothesis, the models (a) and (b), and under the alternative hypothesis, the models (g) and (h). The significant level was $\alpha = 5\%$. The selection of the bandwidths was controlled by selecting the constant C_0 . We notice that the choice of the bandwidth does not have a big impact on the rejection probabilities for all cases. These results reassure that the statistical inference from the proposed test is very stable.

We finally examined the validation of Algorithm 2.2 for surface comparison with spatial correlated errors. We also considered the models (a) and (b) for the null hypothesis, and the models (g) and (h) for the alternative hypothesis. The random errors η_{ij} are generated from a Gaussian process with zero mean and exponential covariance function

$$Cov(\eta_{ij}, \eta_{ik} | \mathbf{X}_{ij}, \mathbf{X}_{ik}) = \sigma^2 \exp \left(- \frac{\|\mathbf{X}_{ij} - \mathbf{X}_{ik}\|}{\varphi} \right), \quad (3.10)$$

with $\sigma^2 = 0.5$ and $\varphi = 0.15$. We observe, in Table 4, that the approximation of level is slightly worse than that in the case of independent errors and the test is also less powerful but still in

a reasonable range. These can be partially explained by a larger bias in estimating regression surface with spatial correlated errors.

3.2. A real data application

We illustrate our testing procedure with real medical rehabilitation image data from a neuromuscular electrical stimulation (NMES) experiment. Pressure ulcers are a major secondary complication of spinal cord injury and can have serious adverse effects on the psychological and physical well-being of the individual. NMES is the application of electrical stimuli to a group of muscles, which is a new clinic tool for pressure ulcers to produce beneficial changes at the user/support system interface by altering the intrinsic characteristics of the user's paralyzed tissue itself (Bogie et al., 2008). In the NMES study from Cleveland FES center (Bogie et al., 2006), pressure intensity data at the seating interface for each patient were recorded as two-dimensional image data with the use of the Tekscan advanced clinseat pressure mapping system (Tekscan, Inc.). The primary goal of the NMES study at Cleveland FES center was to investigate the changes of pressure intensities for each patient before and after NMES treatment.

Figure 1 shows an example of pressure image data from the NMES study. The top panels show the pressure image data before and after treatment for one subject from the NMES study. Both images are on a regular 38×38 grid. The lower panels show the fitted nonparametric local surfaces. A multivariate local linear regression model was considered there. When modeling the data, we transformed the coordinates of the intensities to the region $S = [0, 1] \times [0, 1]$. The estimated optimal bandwidths were obtained by minimizing AIC_C as described in Remark 2.2. The two estimated bandwidth for the two images were approximately equal, $\hat{h}_1 \approx \hat{h}_2 \approx 0.18$. We further explored the empirical varigrams of the residuals of the two images and fit parametric exponential varigram models. The estimates of the correlation parameter ϕ in the exponential covariance function (3.10) were less 0.01 for both images. The results indicate that there is no or very weak spatial correlation for each image.

In order to examine the change of pressure images, we performed the wild bootstrap procedure with the proposed test statistic with a wide range of bandwidths, going from 0.10 to 0.30. The same smoothing parameters were used for each image. The p -values were based on 1000 bootstrap replications. All the results are summarized in Figure 2. The p -value with the optimal bandwidth is extremely small ($< .001$). Across the range of smoothing parameters, the p -values give consistent results. In conclusion, the statistical analysis reflects a strong evidence of a difference between two pressure surfaces (before and after treatment). The NMES does yield effects for the patient with pressure ulcers in this case study.

4. Closing remarks

We studied an $\mathcal{L}^2$ -distance type of test statistic for comparing nonparametric regression surfaces under a completely heteroscedastic nonparametric model. The asymptotic normality of the statistic was also presented. The statistic was extended to be used in the case of spatial correlated errors. Two bootstrap procedures for the test were proposed depending on the nature of random errors.

The numerical results demonstrated that both the wild bootstrap method and the indirect bootstrap method performed excellent even in the case of small samples. When errors were heteroscedastic or sample sizes were small, our method was better than the ANOVA-type test using a χ^2 -approximation by Bowman (2006). Our simulation studies presented above were under a fixed design because the application to the real image data sets was a fixed design. We also performed certain simulations under a random design and studied our method for more than two regression surfaces. We obtained good powers and good approximations of the level

for those cases but did not present the details here. Our two bootstrap procedures proposed could not work in the case of heteroscedastic spatial errors. Heteroscedastic spatial regression models were paid a lot of attention in spatial econometrics (LeSage and Pace, 2009). An appropriate bootstrap method in such a case could be investigated for further research.

Acknowledgments

We are grateful to the three reviewers and the editor for their valuable comments. The research of Xiao-Feng Wang is supported in part by the NIH grant UL1 RR024989.

References

- Bogie KM, Wang XF, Fei B, Sun J. New technique for real-time interface pressure analysis: Getting more out of large image data sets. *Journal of Rehabilitation Research and Development* 2008;45 (4): 523–536. [PubMed: 18712638]
- Bogie KM, Wang XF, Triolo RJ. Long-term prevention of pressure ulcers in high-risk patients: a single case study of the use of gluteal neuromuscular electric stimulation. *Archives of Physical Medicine and Rehabilitation* 2006;87 (4):585–591. [PubMed: 16571402]
- Bowman AW. Comparing nonparametric surfaces. *Statistical Modelling* 2006;6 (4):279–299.
- Cressie, N. *Statistics for Spatial Data*. Wiley; New York: 1993.
- De Jong P. A central-limit-theorem for generalized quadratic-forms. *Probability Theory and Related Fields* 1987;75 (2):261–277.
- Dette H. A consistent test for the functional form of a regression based on a difference of variance estimators. *The Annals of Statistics* 1999;27 (3):1012–1040.
- Dette H, Neumeyer N. Nonparametric analysis of covariance. *Annals of Statistics* 2001;29 (5):1361–1400.
- Francisco-Fernandez M, Opsomer JD. Smoothing parameter selection methods for nonparametric regression with spatially correlated errors. *Canadian Journal of Statistics* 2005;33:279–295.
- Hall P, Hart J. Bootstrap test for difference between means in nonparametric regression. *Journal of the American Statistical Association* 1990;85 (412):1039–1049.
- Hardle W, Mammen E. Semiparametric comparison of regression-curves. *Annals of Statistics* 1990;18 (1):63–89.
- Hardle W, Mammen E. Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 1993;21 (4):1926–1947.
- Hardle, W.; Muller, M.; Sperlich, S.; Werwatz, A. *Nonparametric and Semiparametric Models*. Springer; New York: 2004.
- Hurvich CM, Simonoff JS, Tsai C. Smoothing parameter selection in nonparametric regression using an improved akaike information criterion. *Journal of the Royal Statistical Society, Series B, Methodological* 1998;60:271–293.
- King E, Hart JD, Wehrly TE. Testing the equality of two regression curves using linear smoothers. *Statistics and probability letters* 1991;12 (3):239–247.
- Kulasekera K. Comparison of regression-curves using quasi-residuals. *Journal of the American Statistical Association* 1995;90 (431):1085–1093.
- LeSage, JP.; Pace, RK. *Introduction to Spatial Econometrics*. Chapman and Hall; FL: 2009.
- Neumeyer N, Dette H. Nonparametric comparison of regression curves: An empirical process approach. *Annals of Statistics* 2003;31 (3):880–920.
- Pardo-Fernandez JC. Comparison of error distributions in nonparametric regression. *Statistics and Probability Letters* 2007;77:350–356.
- Pardo-Fernandez JC, Van Keilegom I, Gonzalez-Manteiga W. Testing for the equality of k regression curves. *Statistica Sinica* 2007;17 (3):1115–1137.
- Qiu P. Discontinuous regression surfaces fitting. *Annals of Statistics* 1998;26 (6):2218–2245.
- Takeda H, Farsiu S, Milanfar P. Kernel regression for image processing and reconstruction. *IEEE Transactions On Image Processing* 2007;16 (2):349–366. [PubMed: 17269630]

- Wand, MP.; Jones, MC. Kernel Smoothing. Chapman and Hall; London: 1995.
- Wu C. Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Annals of Statistics* 1986;14:1261–1350.
- Young S, Bowman A. Nonparametric analysis of covariance. *Biometrics* 1995;51 (3):920–931.

Figure 1.

The top panels show the pressure image data before and after treatment for one subject from the NMES study. The lower panels show the fitted nonparametric local surfaces.

Figure 2. The p -values as a function of the bandwidth h obtained from 1,000 bootstrap replications with the proposed testing procedure for the NMES study. The solid horizontal line corresponds to a p -value of 0.05.

Table 1

Simulated level of the tests for various sample sizes and regression functions. T_N denotes the proposed method; T_{ANOVA} denotes the method by Bowman (2006).

		Homoscedastic errors				Heteroscedastic errors			
		T_N	T_{ANOVA}	T_N	T_{ANOVA}	T_N	T_{ANOVA}	T_N	T_{ANOVA}
(n_1, n_2)	Model	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 10\%$
(49,49)	a	0.058	0.099	0.045	0.091	0.046	0.089	0.041	0.079
	b	0.059	0.104	0.023	0.067	0.057	0.108	0.019	0.037
	c	0.047	0.096	0.045	0.089	0.043	0.085	0.035	0.071
	d	0.043	0.091	0.041	0.087	0.060	0.108	0.067	0.119
(100,49)	a	0.053	0.098	0.043	0.090	0.043	0.085	0.045	0.086
	b	0.056	0.106	0.032	0.072	0.058	0.108	0.017	0.038
	c	0.049	0.099	0.041	0.088	0.045	0.089	0.033	0.071
	d	0.051	0.097	0.041	0.087	0.051	0.102	0.064	0.122
(225,225)	a	0.051	0.101	0.045	0.097	0.047	0.093	0.044	0.090
	b	0.052	0.105	0.047	0.098	0.057	0.111	0.031	0.066
	c	0.052	0.098	0.049	0.101	0.046	0.097	0.039	0.082
	d	0.049	0.103	0.048	0.098	0.055	0.108	0.059	0.117

Table 2

Rejection probabilities of the tests under the alternative hypothesis for various sample sizes and regression functions. T_N denotes the proposed method; T_{ANOVA} denotes the method by Bowman (2006).

		Homoscedastic errors				Heteroscedastic errors			
		T_N		T_{ANOVA}		T_N		T_{ANOVA}	
(n_1, n_2)	Model	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 10\%$
(49,49)	e	0.859	0.967	0.712	0.791	0.821	0.885	0.505	0.576
	f	0.574	0.658	0.250	0.375	0.553	0.622	0.131	0.281
	g	0.738	0.829	0.377	0.489	0.698	0.759	0.171	0.274
	h	0.711	0.803	0.318	0.397	0.621	0.699	0.127	0.221
(100,49)	e	0.913	0.954	0.887	0.927	0.844	0.892	0.527	0.599
	f	0.567	0.671	0.191	0.297	0.512	0.617	0.210	0.274
	g	0.795	0.874	0.571	0.663	0.804	0.871	0.319	0.402
	h	0.831	0.905	0.355	0.421	0.794	0.892	0.212	0.389
(225,225)	e	1.000	1.000	1.000	1.000	0.988	1.000	0.826	0.887
	f	0.952	0.983	0.513	0.719	0.901	0.978	0.415	0.570
	g	0.989	1.000	0.831	0.910	0.955	0.981	0.612	0.706
	h	1.000	1.000	0.802	0.887	0.931	0.969	0.771	0.856

The study of the sensitivity of the bandwidth selection: rejection probabilities for various sample sizes and regression functions. The significant level is $\alpha = 5\%$.

Table 3

		Homoscedastic errors				Heteroscedastic errors			
(n_1, n_2)	Model	$C = 0.8$	$C = 1$	$C = 1.2$	$C = 1.5$	$C = 0.8$	$C = 1$	$C = 1.2$	$C = 1.5$
(49,49)	a	0.057	0.052	0.060	0.055	0.051	0.047	0.054	0.059
	b	0.059	0.059	0.051	0.055	0.066	0.059	0.064	0.067
	g	0.715	0.731	0.703	0.742	0.711	0.692	0.733	0.680
	h	0.713	0.729	0.699	0.718	0.667	0.635	0.612	0.710
(100,49)	a	0.054	0.049	0.053	0.059	0.045	0.043	0.052	0.056
	b	0.050	0.055	0.057	0.061	0.043	0.047	0.053	0.059
	g	0.756	0.782	0.806	0.771	0.723	0.801	0.751	0.759
	h	0.812	0.857	0.792	0.823	0.765	0.799	0.810	0.771

Table 4

Rejection probabilities under the null hypothesis for various sample sizes and regression functions. The errors are spatial correlated.

		$n_1 = n_2 = 49$		$n_1 = n_2 = 100$		$n_1 = n_2 = 225$	
Model		$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 10\%$
a		0.041	0.079	0.040	0.081	0.053	0.107
b		0.061	0.116	0.058	0.113	0.055	0.109
g		0.579	0.652	0.704	0.800	0.873	0.925
h		0.617	0.713	0.713	0.834	0.935	0.992