# Comparisons of Two and Three Dimensional Regression Surfaces

K.B. Kulasekera
*Department of Mathematical Sciences*
*Clemson University*
*Clemson, SC 29634*
and
Colin Gallagher
Department of Mathematics and Statistics
CSU, Chico
Chico, CA 95929

**Abstract**

We consider testing the equality of two regression surfaces using two independent samples. A test that is free of the restriction of having identical design points for both samples is proposed. The asymptotic distribution is given and results from a simulation study are presented that show the superior power properties of this test over a competing test in a variety of cases. The method is then applied to a real data set.

AMS Subject Classifications: 62G08, 62G10, 62H15.

Key Words: Kernel estimator; Design variables; Nonparametric test.

# 1. Introduction

In scientific studies, it is very common to compare groups of individuals (or items) in terms of some parameter. For example, a cancer researcher may wish to compare the mean remission time for male and female patients. If the experimenter has auxiliary information such as the age, the blood counts at diagnosis, DNA index etc., of each patient, then he or she may use regression to model the mean remission time as a function of such covariates in each population and compare the two regressions.

We observe data in the form $\{(Y_{1i}, \mathbf{x}_{1i})\}$ $i = 1, ..., n$ and $\{(Y_{2j}, \mathbf{x}_{2j})\}$, $j = 1, ..., m$, with

$$\begin{aligned} Y_{1i} &= f(\mathbf{x}_{1i}) + \epsilon_{1i}, \\ Y_{2j} &= g(\mathbf{x}_{2j}) + \epsilon_{2j}, \end{aligned} \tag{1}$$

where $\epsilon_{1i}, i = 1, ..., n,$ and $\epsilon_{2j}, j = 1, ..., m,$ are i.i.d. random errors for each of the two groups. Here $\mathbf{x}_{1i} = (x_{11i}, ..., x_{p1i})'$ and $\mathbf{x}_{2j} = (x_{12j}, ..., x_{p2j})'$ are the $p$ dimensional covariate measurements ($p = 2, 3$) for the two groups. We are interested in testing the hypothesis

$$H_0 : f = g \quad vs \quad H_1 : f \neq g \tag{2}$$

over the domain of the covariate $\mathbf{x} = (x_1, ..., x_p)'$, which is taken as $[0, 1]^p$, $p = 2$ or 3 in this article. The covariate values (design points) for the two samples may be different. We assume homogeneous error variances, i.e., $Var(\epsilon_{11}) = \sigma_1^2$ and $Var(\epsilon_{21}) = \sigma_2^2$.

Nonparametric testing of the hypotheses of equality of two curves (single covariate) has been discussed by many authors. Among them, Hall and Hart (1990), King, Hart and Wehrly (1991), Delgado (1993), Kulasekera (1995) and Hall, Huber, and Speckman (1997) give methods of testing under a variety of design constraints.

There is very little literature on testing the equality of two surfaces in a completely nonparametric framework even for the two dimensional case. The major difficulty in the testing problem in higher dimensions is efficient estimation of the regression surfaces. The comparison of the volume between the surfaces is very inefficient even in two dimensions. The only literature that is available in testing in a general multidimensional case is for lack of fit testing using a single sample (Hart, 1997). In this framework Hart (1997) proposes a method that transforms a model into a setting $Y = g(t(\mathbf{X})) + \epsilon$ for a user defined function $t(\mathbf{X})$ of the covariates. This makes the problem one dimensional and the wrong specification of $t$ can lead to trivial power. The other method that has been used in multidimensional settings is a modified version of the order selection criteria of Eubank and Hart (1992). However, this method suffers from the curse of dimensionality in that larger dimensions reduce the power significantly in these lack of fit tests.

In this paper we propose test statistics for testing the equality of two regression surfaces $f$ and $g$ (two or three dimensional covariate) that are specified only in terms of some smoothness conditions. In testing the null hypothesis ($H_0 : f = g$) we consider statistics based on estimated projections (into one dimension) of the squared difference $(f - g)^2$ onto a space spanned by a linear combination of the covariates. In this way we avoid the problem of directly estimating the two higher dimensional regression surfaces. We can reasonably expect this method to be more efficient than tests based on the difference of estimated regression functions, i.e., $\hat{f} - \hat{g}$. The treatment of two or three dimensional surfaces does not address the general testing situation. However, even the testing in two dimensions is much more complicated compared with testing the equality of two curves. In what follows, we argue that testing methods for small dimensions would serve experimenters in many situations.

When there are several covariates, experimenters use dimension reduction techniques to reduce the dimension of the covariates without loosing the essential information. Well known dimension reduction techniques include the Sliced Inverse Regression (SIR) (Li, 1991) and Projection Pursuit Regression (PPR) (Friedman and Stuetzle, 1981). Tools of this type have become popular in many data analysis problems with high dimensional covariates. These methods typically produce low dimensional (two or three, (Li, 1991)) regression models. Thus, if one were able to reduce the dimension of the original covariate, the mean functions of the resulting models, which are functions of a small number of new covariates, can be examined in the comparison of the two groups. Here, we describe a method that can handle two or three dimensional surfaces. We believe the method we propose would have good applicability in data analysis subsequent to dimension reduction.

In testing (2), we consider the projection of the squared difference $\psi(x_1, ..., x_p) = (f(x_1, ..., x_p) - g(x_1, ..., x_p))^2$, $p = 2$ or 3, onto a space spanned by a linear combination of the covariates $\mathbf{x}$. For example, for the three variate real valued function $\psi(x_1, x_2, x_3)$, we can define the projection onto $x_1$ as

$$\psi_0(x_1) = \int \psi(x_1, x_2, x_3) dx_2 dx_3.$$

Now, since $\psi \geq 0$ for all values of $\mathbf{x} = (x_1, x_2, x_3)'$, $\psi_0 = 0$ iff $\psi = 0$ for continuous functions $f$ and $g$. Thus, we can test

$$H_0 : \psi_0 = 0 \quad vs \quad H_1 : \psi_0 > 0 \tag{3}$$

for all values of $x_1$ instead of the original hypotheses (2).

If the two sample designs are identical,($\mathbf{x}_{1i} = \mathbf{x}_{2i} = \mathbf{x}_i, i = 1, ..., n$), where $\mathbf{x}_i$ are the common design points, the differences of the responses $Y_{1i} - Y_{2i} = D_i$, have means $f(\mathbf{x}_i) - g(\mathbf{x}_i), i = 1, ..., n$. Then, we develop estimators of $\psi_0(x_{1i}), i = 1, ..., n$ based on the differences $D_i, i = 1, ..., n$. These estimators can be combined to form a suitable test statistic, $\sum_{i=1}^n \hat{\psi}_0(x_{1i})$, where $\hat{\psi}_0(x_{1i})$ is the estimator for $\psi_0(x_{1i})$. This statistic is normalized so that it can detect departures from the null hypothesis in (3).

For non-identical designs (same sample sizes), the test can be developed under the condition that in large samples, for every point in the first design, there would be an arbitrarily close point in the second design. This is a generalization of the matching condition used by Hall and Hart (1990) to a multi-dimensional case using a suitable distance measure such as the Euclidean distance on the covariate domain.

Subsequently, the test is extended for cases with unequal sample sizes using a simple method to create pseudo observations, thus compensating for the smaller number of observations in one of the samples. It is shown that asymptotic properties of our test remain the same with the proposed generation of the pseudo values.

The proposed test can be extended in a couple of ways. Results obtained here for $[0,1]^p$ can be easily extended to other types of two or three dimensional rectangles by suitable transformations. In this article we describe the procedure for non-random covariate vectors. One may be able to handle the case with random covariates in a similar manner under suitable restrictions on the conditional distribution of the responses given the covariates. However, verification of such conditions for the conditional distributions (and the joint distribution of the covariates) may be difficult. We have no results in this regard at this time.

We compare this method to the Kolmogorov-Smirnov type test introduced by Delgado (1993) which can be used with common designs and extended to unequal covariate designs in the same manner as our test. Simulations show that for many functions $f$ and $g$, the power of that test is significantly smaller compared with the projection method proposed here, the lack of power becoming worse when the two designs are not identical.

The article is organized in the following manner. Section 2 gives details of the procedure including asymptotic results. Some simulation results and an example with real data are given in Section 3. Section 4 gives a few concluding remarks and finally in the Appendix we sketch the proofs.

# 2. Test Statistics and Properties

We first describe the case with equal sample sizes for the two groups. This will be extended to non-identical designs with the same sample sizes and then to unequal designs. We use the term $p$ dimensional covariate to refer to either the two or three dimensional case. We begin by listing some assumptions that will be used in the sequel.

$\mathcal{A}_1$**:** design densities for the two samples are the same and the design points $\mathbf{x}_{1i}$ and $\mathbf{x}_{2j}$ become dense everywhere in the covariate domain $[0,1]^p$ as $m, n \to \infty$.

$\mathcal{A}_3$**:** functions $f$ and $g$ are twice continuously differentiable in all arguments over the covariate domain.

## 2.1 Identical Designs Case

We first decribe the procedure for identical covariate designs. Suppose the common covariate values in the samples are $\mathbf{x}_i = (x_{1i}, ..., x_{pi})', i = 1, ..., n$ for a $p$ dimensional covariate $\mathbf{x} = (x_1, ..., x_p)'$ and the two samples follow the models

$$\begin{aligned} Y_{1i} &= f(\mathbf{x}_i) + \epsilon_{1i}, \\ Y_{2i} &= g(\mathbf{x}_i) + \epsilon_{2i}. \end{aligned} \tag{4}$$

It is further assumed that within each group the errors are identically distributed but the distributions of the $\epsilon_1$'s and $\epsilon_2$'s may be different. It is also assumed that $E(\epsilon_{11}^{2l+1}) = E(\epsilon_{21}^{2l+1}) = 0, l = 0, 1$ and $E(\epsilon_{11}^4)$ and $E(\epsilon_{21}^4)$ are both finite. We have

$$D_i = f(\mathbf{x}_i) - g(\mathbf{x}_i) + \epsilon_{1i} - \epsilon_{2i}, i = 1, ..., n$$

where $D_i = Y_{1i} - Y_{2i}$ as defined in the Section 1 above. Then $Var(\epsilon_{1i} - \epsilon_{2i}) = \sigma_1^2 + \sigma_2^2 = \sigma^2$, (say). This leads to

$$\begin{aligned} D_i^2 &= (f(\mathbf{x}_i) - g(\mathbf{x}_i))^2 + 2(f(\mathbf{x}_i) - g(\mathbf{x}_i))(\epsilon_{1i} - \epsilon_{2i}) + (\epsilon_{1i} - \epsilon_{2i})^2 \\ &= \psi(\mathbf{x}_i) + \eta_{1i}, \end{aligned}$$

where $\psi(\mathbf{x}_i)$ is as defined in the previous section and $\eta_{1i} = 2(f(\mathbf{x}_i) - g(\mathbf{x}_i))(\epsilon_{1i} - \epsilon_{2i}) + (\epsilon_{1i} - \epsilon_{2i})^2$. Thus,

$$D_i^2 - \sigma^2 = \psi(\mathbf{x}_i) + \eta_i, \tag{5}$$

where $\eta_i = \eta_{1i} - \sigma^2$, $i = 1, ..., n$ and, $E(\eta_i) = 0, i = 1, ..., n$. It should be noted that the errors $\eta_i$, $i = 1, ..., n$ do not have equal variances unless $f = g$. Define a linear combination of the covariates $x_1, ..., x_p$ by $u = \mathbf{c}'\mathbf{x}$, where $\mathbf{c} = (c_1, ..., c_p)'$. Then the sample values of $u$ are defined as $u_i = \mathbf{c}'\mathbf{x}_i, i = 1, ..., n$. Now, consider the integral of the curve that results by intersecting the surface $\psi(\mathbf{x})$ with the plane through $u = \mathbf{c}'\mathbf{x}$ defined as

$$\psi_0(u) = \int_{\mathbf{c}'\mathbf{x}=u} \psi(\mathbf{x})|ds|$$

where $|ds|$ is a small increment in the direction of $u = \mathbf{c}'\mathbf{x}$. For example, in the two dimensional case with $u = c_1x_1 + c_2x_2$, one can write

$$\psi_0(u) = \int \psi((u - c_2x_2)/c_1, x_2)\frac{\sqrt{c_1^2 + c_2^2}}{|c_1|}dx_2,$$

and for $c_1 = 1, c_2 = 0$, we get the projection of $\psi$ onto $x_1$. It should be noted that the limits of the integration may depend on the values of $u$. The resulting function $\psi_0$ is twice continuously differentiable for all values of $u$ when the projections are onto the $x_j$'s, $j = 1, .., p$. For all other projections, $\psi_0$ is piecewise twice continuously differentiable. $\psi_0$ can be thought of as the average of the $\psi$ values along the line defined by $u = \mathbf{c}'\mathbf{x}$ for a given value of $u$. If one were only interested in projecting onto a covariate axis, then an estimate of $\int \psi(\mathbf{x})d\mathbf{x}$ could be used as a test statistic. However, the projection onto a different plane may provide a better power for some functions $f - g$ when the sample sizes are not very large. Thus we develop the method in a more general form.

If we knew $\sigma^2$, we could estimate $\psi_0$ using a linear smoother on $D_i^2 - \sigma^2$ against $u_i, i = 1, ..., n$. This would amount to a scatter plot smoothing when the noise has a non-constant variance. In a sense, this parallels the integration of multivariate smoothers using product kernels to extract additive components in a Generalized Additive Model (GAM) (Linton and Härdle, 1996). In Linton and Härdle (1996), additive components of the GAM were looked at as projections of the additive function onto each covariate alone. Here, model (5) is not necessarily additive, but we are considering the projection of the function $\psi$, which is a regression function in some sense, onto a single covariate (or a linear combination of the $p$ predictors).

Since $\sigma^2$ is assumed unknown, we use an estimator $\hat{\sigma^2}$ in place of of $\sigma^2$ in the smoothing of $D_i^2 - \sigma^2$ against the $u_i$'s.. Here we propose to use $\hat{\sigma^2} = \mathbf{D}'G\mathbf{D}$ where $\mathbf{D} = (D_1, ..., D_n)'$, and $G$ is the matrix of the quadratic form for an estimator of the error variance in a nonparametric regression model. If we choose $G$ to be the matrix for a difference estimator (Hall, Kay and Titterington, 1990), then the corresponding estimator $\hat{\sigma^2}$ is unbiased for $\sigma^2$ under the null hypothesis $f = g$. Even if $f \neq g$, such difference estimators would be consistent for the variance $\sigma^2$ when $p = 1$ and the covariates are ordered. However, when $p > 1$, one cannot uniquely order the covariates and therefore consistent estimation of $\sigma^2$ has to be done using a systematic grouping of the $D_i$'s. As shown in Section 2.4, we can rearrange the covariates in a way that the difference estimator of the variance would be consistent for $\sigma^2$ under both the null and the alternative hypothesis. Estimation of the variance in nonparametric regression with multiple covariates has has not been extensively discussed in the literature. Dette and Derbort (2001) have given an ad hoc estimator that may work in some situations. The method we propose gives a consistent estimator for the error variance in nonparametric regression models with $p = 2$ and $p = 3$. In fact the same idea can be extended to variance estimation in higher dimensions (Kulasekera and Gallagher, 2001).

For convenience we use a variance estimator based on the first order differences of the responses (rearranged as described in Section 2.4, if necessary), where the corresponding $G$ is given by

$$G = \frac{1}{2(n-1)} \begin{pmatrix} 1 & -1 & 0 & & . & . & . & 0 \\ -1 & 2 & -1 & 0 & . & . & . & 0 \\ 0 & -1 & 2 & -1 & 0 & . & . & 0 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 0 & . & . & . & 0 & -1 & 2 & -1 \\ 0 & . & . & . & . & 0 & -1 & 1 \end{pmatrix}_{n \times n} .$$

Then, an estimator for $\psi_0$ can be written as

$$\hat{\psi}_0(u) = \sum_{i=1}^{n} w_i(u)[D_i^2 - \mathbf{D}'G\mathbf{D}] \tag{6}$$

$$= \mathbf{D}'[W(u) - G]\mathbf{D} \tag{7}$$

where $W(u) = diag(w_1(u), ..., w_n(u))$ and $w_i(u)$ are the weights for a linear smoother which we choose so that they sum to 1. For example we may use a local linear smoother (Fan, 1992) or the Nadaraya-Watson type smoother(Nadaraya, 1964). In case of the latter, the weights are given by

$$w_i(u) = \frac{K\left(h^{-1}(u - u_i)\right)}{\sum_{j=1}^{n} K\left(h^{-1}(u - u_j)\right)}$$

where $K$ is a kernel function and $h$ is a smoothing parameter. Typically, the function $K$ is chosen to be a probability density function symmetric about 0 and the smoothing parameter $h$ is selected to satisfy the following assumption.

$\mathcal{A}_3$**:** The bandwidth sequence $h$ satisfies $nh \to \infty$ and $nh^5 \to 0$ as $n \to \infty$.

The null hypothesis in (2) is rejected for "large" values of $T = \sum_{i=1}^{n} \hat{\psi}_0(u_i)$. We obtain a formal test using the large sample properties of a properly normalized version of $T$. For Nadaraya-Watson linear estimators with uniform kernel functions we have the following theorem.

**Theorem 2.1** *Under the assumptions* $\mathcal{A}_1$,$\mathcal{A}_2$ *and* $\mathcal{A}_3$, *and the null hypothesis,* $T_0 = T/\sqrt{n\hat{\sigma}^4}$ *converges in distribution to a standard normal variable as* $n \to \infty$. *Then, a size* $\alpha$ *test for testing (2) is given by the rule: "Reject* $H_0$ *if* $T_0 > z_\alpha$, *where* $z_\alpha$ *is the upper* $\alpha$ *quantile of a standard normal variable".*

**Proof:**

In the appendix, we give the proof of the theorem for projections onto $x_1$ when the $x_1$ coordinates are equi spaced. For unequal spacings and other projections the proof can be extended with minor modifications assuming that the design points become dense inside the design space when $n \to \infty$.

**Remark 2.1** In the proposed procedure smoothing is done after taking the differences. This is a smoothed von Neumann type test that essentially compares two estimators of varaiance under the null hypothesis. An alternative could be to smooth $Y$'s against $u$'s in the two samples and examine the difference between those smooths. However, such a test actually tests the equality of the functions $\tilde{f}$ and $\tilde{g}$ where $Y_{1i} = \tilde{f}(\mathbf{c}'\mathbf{x}_{1i}) + \tilde{\epsilon}_{1i}$ and $Y_{2j} = \tilde{g}(\mathbf{c}'\mathbf{x}_{2j}) + \tilde{\epsilon}_{2j}$ for some noise stucture $\tilde{\epsilon}$. This is somewhat in the sipirt of lack of fit tests proposed by Hart (1997). However, there is no clear idea how to pick $\mathbf{c}$ so that one would get a non trivial power.

**Remark 2.2** Under the alternative hypothesis,

$$E(T_0) \approx \frac{\sum_{i=1}^{n} \psi(\mathbf{x}_i)}{\sqrt{n}},$$

which diverges to $\infty$ at a rate of $\sqrt{n}$. Thus the proposed test detects local alternatives of the type $f(\mathbf{x}) - g(\mathbf{x}) = n^{-1/4} g_0(\mathbf{x})$ with non trivial power..

## 2.2 Non-Identical Designs with $m = n$

When the two covariate designs are not the same (but with same number of design points), we follow a method that is similar to the matching technique used by Hall and Hart (1990). Suppose the data take the form $(Y_{1i}, \mathbf{x}_{1i}), (Y_{2i}, \mathbf{x}_{2i}), i = 1, ..., n$ with $Y_{1i} = f(\mathbf{x}_{1i}) + \epsilon_{1i}$ and $Y_{2i} = g(\mathbf{x}_{2i}) + \epsilon_{2i}$. We assume that for every point $\mathbf{x}_{1i}$ in the first design, there is a point in the second design $\mathbf{x}_{2i'}$ such that $\max_i \| \mathbf{x}_{1i} - \mathbf{x}_{2i'} \| = O(a_n)$, $a_n \to 0$, as $n \to \infty$ for a suitable sequence $a_n$. In practice, for each design point $\mathbf{x}_{1i}$ for the first sample, we can pick the closest design point $\mathbf{x}_{2i'}$ (with respect to the Euclidean distance) from the second sample. If the closest point has been already picked, then we propose to choose the next closest point. Then, we match the responses corresponding to these two design points to get $D_i = Y_{1i} - Y_{2i'}, i = 1, ..., n$, and set $u_i = \mathbf{c}'\mathbf{x}_{1i}, i = 1, ..., n$. We can then write the initial model as

$$D_i = f(\mathbf{x}_{1i}) - g(\mathbf{x}_{1i}) + g(\mathbf{x}_{1i}) - g(\mathbf{x}_{2i'}) + \epsilon_{1i} - \epsilon_{2i'}, i = 1, ..., n.$$

In this matching, under the hypothesis that $f = g$, it can be shown that the statistic

$$T_0 = \frac{\sum_{i=1}^{n} \hat{\psi}_0(u_i)}{\sqrt{n\hat{\sigma}^4}}$$

has no effect from the terms corresponding to $g(\mathbf{x}_{1i}) - g(\mathbf{x}_{2i'}), i = 1, ..., n$ as $n \to \infty$ provided $\sqrt{n}a_n^2 \to 0$ as $n \to \infty$. This is a reasonable assumption for the covariate designs when $p \leq 3$. With these assumptions, the asymptotic properties of the test statistic $T_0$ under the null hypothesis will remain the same as in Theorem 2.1.

## 2.3 Unequal Sample Sizes

In this section, we discuss possible extensions of the above test to situations with unequal sample sizes as given in (1). Without any loss of generality, we will assume that $m < n$ and that the $m$ design points for the second group are the same as the first $m$ design points for the first group. If not, one can match the $m$ covariate values from the second sample with $m$ covaraite values from the first sample as above and consider those as 'common' design points. Thus, in this situation we have $n - m$ more design points in the first sample than the second. Suppose $u_i, i = 1, ..., n$ are the projected covariate values for the first design. Again, without loss of generality we may take $u_i, i = 1, .., m$ as the ordered $u$ values corresponding to the common $m$ design points for the two samples and, $D_i, i = 1, ..., m$ to be the corresponding differences, $D_i = Y_{1i} - Y_{2i}$. Our proposal is to generate $n - m$ pseudo-differences $\tilde{D}_j$ at $u_j, j = m+1, ..., n$. We can generate pseudo-differences using one of the following schemes.

**Scheme 1:** Find an integer $i, i = 1, ..., m$ and pick an integer $l > 0$ such that $u_1 \leq u_{i-l+1} < u_j < u_{i+l} \leq u_m$ for each $j = m + 1, ..., n$. Now set $\tilde{D}_j = \sum_{k=i-l+1}^{i+l} a_k D_k$, $j = m + 1, ..., n$. where the coefficients $a_k$ are chosen such that $\sum a_k = 1$. If a $u_j, j = m+1, ..., n$ falls below $u_1$, set $\tilde{D}_j = D_1$ and for $u_j > u_m, j = m + 1, ..., n$ set $\tilde{D}_j = D_m$.

**Scheme 2:** Consider a design point $\mathbf{x}_{1j}, j = m + 1, .., n$ from the first sample that would not be matched with a design point from the second sample. Select the design point $\mathbf{x}_{1i}, i = 1, ..., m$ that is closest to this point $(\mathbf{x}_{1j})$. Now define $\tilde{D}_j = Y_{1j} - Y_{2i}$ where $Y_{1j}$ corresponds to $\mathbf{x}_{1j}$ and $Y_{2i}$ is the response from the second sample that corresponds to $\mathbf{x}_{1i}$.

We combine the two sets, $D_i, i = 1, ..., m$ and $\tilde{D}_j, j = m + 1, ..., n$ to get $n$ differences at $u_i, i = 1, ..., n$. Then the tests statistic $T_0$ is formed using these observations. The $\tilde{D}$'s will also have zero mean under the null hypothesis under both generation schemes. The differences resulting from the first scheme will have different variances compared with the original $D$'s. However, it is seen that the sampling distribution of $T_0$ remains the same as in Theorem 2.1 under some reasonable assumptions. In particular, for both pseudo difference schemes, we have

**Theorem 2.2** *Under the null hypothesis and under the assumptions, $\frac{n}{m} \to 1$, and $\frac{n-m}{\sqrt{m}} \to 0$, as $m \to \infty$, $T_0$ in Theorem 2.1 converges in distribution to a standard normal variable.*

**Proof:**

The proof follows on the same lines as in the proof of Theorem 2.1 once we make a few minor adjustments. These are given in the appendix.

**Remark 2.3** It should be noted that the test proposed by Delgado (1993) can also be modified to test (2) by matching the responses in the above manner. The large sample properties of that test will also hold under the assumptions of this article. However, our simulations show that the projection method is much superior in power compared with Delagado's test for most alternatives, while they both enjoy similar size properties.

## 2.4 Variance Estimator

We discuss the development of a variance estimator for multiple covariate case here. This approach applies to any nonparametric regression model with two or three predictors. Thus, in this formulation we use a general model using a slightly different notation. For an estimator of the error variance in general non parametric regression models reader is referred to Kulasekera and Gallagher (2001).

Supposes we have data in the form $\{(Z_i, \mathbf{t}_i), i = 1, ..., n\}$ following a model

$$Z_i = \mu(\mathbf{t}_i) + \theta_i,$$

where $\theta_i$'s are *iid* with zero mean and finite variance $\gamma^2$. We assume $\mu$ is twice continuously differentiable in all arguments (including mixed derivatives) and the non-random covariate $\mathbf{t} = (t_1, ..., t_p)'$ takes values in $[0, 1]^p$, $1 \leq p \leq 3$. In the one dimension case, without loss of generality, we can assume that the covariates are ordered. Then, for example, for the estimator $\hat{\gamma}^2$ with first order differences (Hall, Kay and Titterington, 1990),

$$\hat{\gamma}^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} [Z_{i+1} - Z_i]^2, \tag{8}$$

the consistency (and asymptotic normality ) follows from the fact that successive differences $\mu(t_{1i+1}) - \mu(t_{1i})$ have negligible effect when the regression function $\mu$ is continuously differentiable and the covariate values become dense in $[0,1]$ as the sample size approaches $\infty$. Further, if $\mu \equiv 0$, the difference estimator is unbiased for $\gamma^2$.

If $p > 1$, unique ordering the covariates is not possible. Thus, the contribution from terms such as $\mu(\mathbf{t}_{i+1}) - \mu(\mathbf{t}_i)$ for a difference estimator of type (8) is not necessarily negligible for all permutations of the responses $Z_i, i = 1, ..., n$. For this situation, we propose a re-arrangement of covariates in which we can form successive differences $Z_{i+1} - Z_i$ in such a way that the effect of terms like $\mu(\mathbf{t}_{i+1}) - \mu(\mathbf{t}_i)$ in an estimator of the error variance $\gamma^2$ becomes negligible when the sample size increases. This rearrangement would ensure that the difference estimator of the error variance is consistent and asymptotically normal. For ease in explanation, we consider first order difference estimators of type (8).

First consider the case $p = 2$. Then, our covariates are $t_1$ and $t_2$. Now, we permute all the data values according to the increasing order of the covariate $t_1$. Then, we group the $t_1$ values into $m+1$ groups using a partition $0 = a_0 < a_1 < a_2 < ... < a_m < a_{m+1} = 1$. Then we permute the responses within each slice $I_i = [a_i, a_{i+1}], i = 0, ..., m$ according to the increasing order of $t_2$ values. Now, we take successive differences of the permuted responses in the following order: select and label the response corresponding to the smallest $t_1$ value and the smallest $t_2$ value in slice $I_0$ as $Z_1$ and the response with the next smallest $t_2$ within the slice $I_0$ as $Z_2$. Likewise, the response $Z_3$ would be the response that corresponds to the next smallest $t_2$ within the slice $I_0$ and so on. Once we reach the response with the largest $t_2$ value in $I_0$, the next response would be the one that corresponds to the largest $t_2$ value in $I_1$. Then we go down the $I_1$ slice according to the value of $t_2$ and, once we pick the response attached to the smallest $t_2$ value in $I_1$, we move to the response that corresponds to the smallest $t_2$ value in the slice $I_2$. The process continues until all responses have been accounted for. The rearranged responses are then used in a difference estimator. We pick the $m$ slices in such a way that $m \to \infty$ as $n \to \infty$ and all the slice sizes decrease to zero at the same rate while the number of the responses within a slice go to $\infty$. For example, we may take $m = n^{1/2}$. Thus, if $(t_1, t_2)$ $(t_1', t_2')$ are two successive points in forming the differences, $\| \mu(t_1, t_2) - \mu(t_1', t_2') \| \to 0$ as $n \to \infty$ due to the fact that covariate values become dense within the domain $[0,1]^2$ and due to the smoothness of $\mu$ with respect to all its arguments. The asymptotic properties of the estimator remain the same as in the case $p = 1$.

If $p = 3$, the slicing is done on the covariates $(t_1, t_2)$ to form $m$ two-dimensional rectangles and we take the successive differences moving up and down according to the values of the $t_3$ coordinate within adjacent rectangles. The consistency of the difference estimator follows if the selected size of each rectangle is small. In particular, if for all points $\mathbf{t}_1, \mathbf{t}_2$ in a two-dimensional rectangle, $\sup_{\mathbf{t}_1,\mathbf{t}_2} \| \mathbf{t}_1 - \mathbf{t}_2 \| \to 0$ as $n \to \infty$ where $\| \,.\, \|$ is the Euclidean norm in the plane, then the smoothness of the function $\mu$ gives that the contribution of the differences of successive function values is asymptotically negligible in the expression for the variance estimator. Such difference estimators would retain the asymptotic properties of the estimator with $p = 1$ when the number of rectangles tends to $\infty$ and the covariates become dense in $[0, 1]^p$ as $n \to \infty$. We examine the performance of estimators of this type through a simulation study for $p = 2$ and $p = 3$ in Section 3.

# 3. Empirical Results and an Example

In this section we provide results of a small simulation study and present a real data example.

## 3.5 Simulations

We conducted a simulation study to assess the power (and size) properties of the proposed projection method. In the simulations we used two and three predictors with several regression functions; functions ranging from simple additive models of polynomials to multiplicative regression functions of periodic functions. The error distribution was taken to be normal and $t_6$ for the two populations where the covariate designs were taken to be non-identical with equal sample sizes.

The two samples were matched for "close" design points in the following manner: first generate the design points for the first group where each covariate value is generated by a uniform random variable in $[0, 1]$. Then, perturb each coordinate of those design points by a fixed amount, say $\pm\delta$ and take those points as the design points for the second sample. For example, for $p = 2$, if $(x_{11i}, x_{21i})$ is the $i$th design point for the first sample we use $(x_{11i} \pm \delta, x_{21i} \pm \delta)$ as the closest design point from the second sample and the match the corresponding responses to get the difference $D_i, i = 1, ., n$. This selection may not necessarily match the closest points from the two groups with respect to the Euclidean distance, but we used several choices of $\delta$ values and $x_{j1} \pm \delta, j = 1, ..., p$ for each covariate, moving each covariate value both up and down to get a reasonable feel for the closest matching scenario. If any $x_{j1} \pm \delta, j = 1, ..., p$ value falls below 0 or above 1 we replace that value by 0 or 1 respectively. The values of $\delta$ used in the simulations were

$0.01, 0.025$ and $0.05$. The results we present in the tables correspond to $\delta = 0.01$ for both $p = 2\&3$, i.e. $x_{j2} = x_{j1} + (-1)^j \delta, j = 1, 2, 3$. We used the projection to each axis ($u = x_j, j = 1, 2, 3$) , and the projection $u = \sum_{j=1}^{p} x_j$ in calculating the power and the size of the tests.

The sample size from each population was taken to be 50 for the two predictor case. In the case of three predictors, we used sample sizes of 64 from each population. In the two covariate case, we also conducted a few simulations with sample size 100 from each population to examine the convergence rate of power and size of the projection method.

The results we present here are with standard normal and $t_6$ errors for the two groups. In this choice of normal error, the signal to noise ratio is rather low for some regression functions. In such cases, we also tried using a smaller error variance (0.01) and the power properties of the projection method improved dramatically.

The number of simulations was 1000 where Splus was used as the programming language. Tables 1-6 give the results of this empirical study. We also calculated the power and the size for the test by Delgado (1993) (given as DT in the tables) for comparison.

In all the simulations, the variance was estimated using the slicing method we have described in Section 2.4. For the two covariate case we divided the $x_1$ axis into 10 equal intervals while in the three covariate case we divided the $x_1, x_2$ plane into 16 squares in implementing the estimation technique. To study the sampling properties of the variance estimator, we generated several multiple regression models of type

$$Z = \mu(\mathbf{t}) + \theta,$$

where $\theta$ is a standard normal variable, $\mathbf{t} = (t_1, ..., t_p)'$ is the vector of covariates from $[0, 1]^p$, $p = 2, 3$, and $\mu$ is the regression function, and applied the proposed variance estimation method. We used sample sizes $25, 64$ and $100$ for two covariates case (dividing the $x_1$ axis into $\sqrt{n}$ intervals) and sample sizes $27, 64$ and $125$ for three covariates case (dividing the $x_1, x_2$ plane into $n^{2/3}$ squares). The average and the standard deviation for variance estimators for 1000 simulations are presented in Table 8.

The examination of the Tables 2-7 show that the projection method has very good size properties. The sample sizes (50 and 64 ) are very moderate (even small) in a study of two or three dimensional regression problems. Even with such sample sizes, the size properties are very satisfactory. When the error distribution is $t$ it appears that there is an increase in the empirical size in some cases although it is not drastic. The size seems to be very similar for all projections for a fixed function and bandwidth seems to have a somewhat visible influence in a few cases. This is not at all surprising. As for power, the projection method seems to give reasonable power in almost all cases we had examined. In almost all cases where the models have high periodic type functions, the projection method is very much superior to the covariate matched Delgado procedure. The only time Delagado method had better power was when the regression functions were polynomial type. This may be due to the fact that in oscillating functions, the partial sums that are the building blocks for the Kolmogorov-Smirnov type Delagado test tend to be small. All the projections we examined gave very close power results in most cases and the effect of the bandwidth was not quite clear.

## 3.6 An Example

In a clinical trial, 59 epilepsy patients suffering from simple or complex partial seizures were randomized to two groups (Thall and Vail, 1990). The first group with 28 patients received a placebo while the other 31 patients received the anti-epileptic drug Progabide as an adjutant to standard chemotherapy. The number of seizures was counted over four two-week periods, denoted by $Y_1, Y_2, Y_3$ and $Y_4$, respectively. The two covariates in each group were a baseline seizure rate which was recorded for each patient based on the eight-week prerandomization seizure count and the patient's age. In this example, we are interested in testing the treatment effect using the method we have developed. Thall and Vail (1990) in their analysis note a treatment effect. They also note that Progabide may be contraindicated for patients with high baseline counts.

For technical simplicity, we use the average of the four seizures $Y_1, Y_2, Y_3$ and $Y_4$, denoted by $Y$. We assume that the model $Y = f(x_1, x_2) + \epsilon_1$ holds for the group with the placebo and the model $Y = g(x_1, x_2) + \epsilon$ for the treatment group where, $x_1$ is the age and $x_2$ is the baseline rate. Figure 1 presents the two dimensional kernel smoothing of $f(x_1, x_2)$ for the patients with a placebo while Figure 2 gives a kernel estimate of $g(x_1, x_2)$ for the patients with Progabide treatment. Thall and Vail (1990) noticed that there is an outlier in the Progabide group. In our calculations we omitted that outlier, resulting in two samples of sizes 28 and 30. Since the covariate values were not identical for the 28 cases in the two samples, we first matched them in the following manner: first we applied the rearrangement mechanism for estimation of variance discussed in Section 2.4 to each group of data. Here we divided the $x_1$ axis into seven slices with four responses each. Then we matched the responses corresponding to the $x_2$ values (in the increasing order) within each slice in the two groups. Then we used the scheme 1 in section 2.3 to generate the two needed pseudo differences.

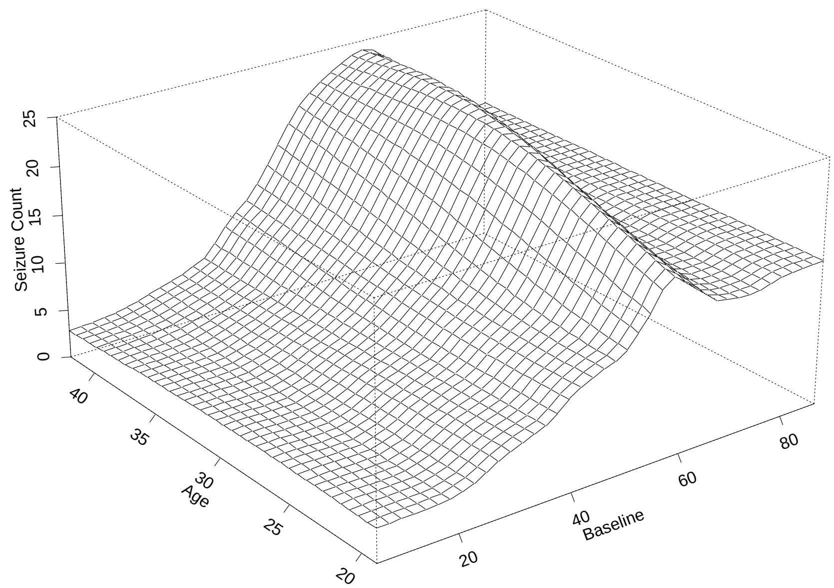


Figure 1: Seizure Count per week without Progabide

Table 1 gives a set of projections and the $p$-values for the test of no treatment effect ($f = g$) using several bandwidths for each projection. It is clear that the bandwidth and the projection both have an effect on the decision. Delgado's (1993) method rejected the null hypotheses of no difference at the 0.05 level.

# 4. Conclusions

In this paper, we have proposed a test statistic for testing the hypothesis of the equality of two regression surfaces. We discussed the asymptotic behavior of the test statistic and our empirical study shows that the proposed testing procedure has good power and size properties.

The test statistic above involves a smoothing parameter, and it is clear that the choice of smoothing parameters play a critical role in the performance of the proposed statistics. The use of optimal smoothing parameters would enhance the power performance of tests (Kulasekera and Wang, 1997). In our presentation, we used deterministic bandwidths for technical and computational simplicity. However, data-based smoothing parameters would be more suitable. We beleive that the smoothing parameter selection procedures given in Kulasekera and Wang

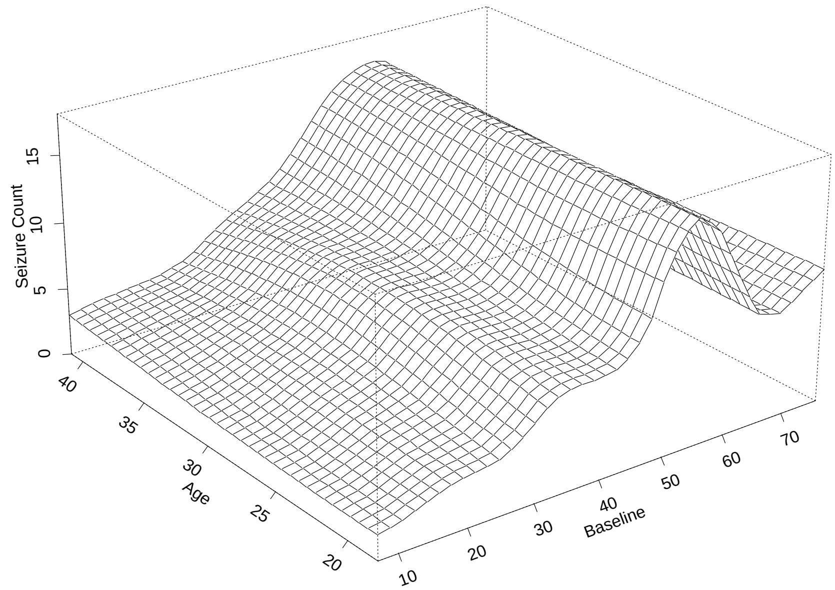


Figure 2: Seizure Count per week with Progabide

(1997) can be applied to the data once the variable $u$ is defined. Also, the choices of the projections were arbitrary in this paper. As long as the squared difference $(f-g)^2$ is non-zero everywhere in the covariate domain, we can anticipate reasonable power for almost all projections. Finally, and most importantly, the development of a test for any dimension remains an open problem.

# 5. Appendix

We give a proof of the Theorem 2.1 for projections onto $x_1$ when the $x_1$ values are equi spaced, using the Nadaraya-Watson (NW, hereafter) smoother with

$$K(x) = 0.5 I_{([-1,1])}.$$

For unequal spacings, for other projections and for other kernel functions the proof can be extended with minor modifications assuming that the design points become dense inside the design space when $n \to \infty$. We use the notation $E_0$ and $Var_0$ to indicate the mean and variance of any statistic under the null hypothesis $f = g$.

First notice that, $E_0(\hat{\psi}_0(x_{1i})) = 0$. Then, consider

$$T = \sum_{i=1}^{n-1} \hat{\psi}_0(x_{1i})$$

Table 1: p values for Epilepsy Seizures Example

| | p-value | | |
|---|---|---|---|
| Bandwidth | $u = x_1$ | $u = x_2$ | $u = (x_1 + x_2)/2$ |
| 3 | 0.053 | 0.033 | 0.048 |
| 5 | 0.045 | 0.022 | 0.054 |
| 7 | 0.069 | 0.079 | 0.050 |
| 10 | 0.119 | 0.112 | 0.037 |

$$= \sum_{i=1}^{n-1} \mathbf{D}'[W(x_{1i}) - G]\mathbf{D}. \tag{9}$$

We include only the sum of the first $n-1$ terms in the sum for $T$, which does not change the asymptotics. Now, let $W^* = \sum_{i=1}^{n-1}[W(x_{1i}) - G]$. Note that $W^*$ is a tri-diagonal matrix of the form

$$W^* = \begin{pmatrix} (w_1 - 1/2) & -1/2 & 0 & & . & . & . & 0 \\ -1/2 & (w_2 - 1) & -1/2 & 0 & . & . & . & 0 \\ 0 & -1/2 & (w_3 - 1) & -1/2 & 0 & . & . & 0 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 0 & . & . & . & 0 & -1/2 & (w_{n-1} - 1) & -1/2 \\ 0 & . & . & . & . & 0 & -1/2 & (w_n - 1/2) \end{pmatrix},$$

where $w_j = \sum_{i=1}^{n-1} w_j(x_{1i})$. For the NW estimator $\hat{\psi}_0$,

$$w_j = \sum_{i=1}^{n-1} \frac{K\left(h^{-1}(x_{1j} - x_{1i})\right)}{\sum_{l=1}^{n} K\left(h^{-1}(x_{1i} - x_{1l})\right)}.$$

If $h$ is the bandwidth used in NW smoother, due to the equal space design assumption, we get $w_j = 1$ for $2nh \le j \le n - 2nh$. Let $k = [2nh]$. By a simple counting argument, we can show that $0 \le |w_j - 1| \le 1$ for $1 \le j \le k$ and $n - k \le j \le n$. Now, partition $W^*$ as (assuming that $n$ is large enough to have a $k$ that makes the formation of the matrices to be meaningful)

$$W^* = \begin{pmatrix} A_{k\times k} & \mathbf{0}_1 & \mathbf{0} \\ \mathbf{0}_2 & W_0^* & \mathbf{0}_3 \\ \mathbf{0} & \mathbf{0}_4 & B_{k\times k} \end{pmatrix}.$$

where

$$\mathbf{0}_1 = \begin{pmatrix} 0 & 0 & . & . & . & 0 \\ 0 & 0 & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & 0 \\ -1/2 & 0 & . & . & . & 0 \end{pmatrix}_{k\times(n-2k)},$$

$$\mathbf{0}_3 = \begin{pmatrix} 0 & 0 & . & . & . & 0 \\ 0 & 0 & . & . & . & 0 \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & 0 \\ -1/2 & 0 & . & . & . & 0 \end{pmatrix}_{(n-2k)\times k},$$

$\mathbf{0}_2 = \mathbf{0}_1'$, $\mathbf{0}_4 = \mathbf{0}_3'$ and $\mathbf{0}$'s are zero matrices with matching sizes. Then, we can rewrite $\mathbf{D}'W^*\mathbf{D}$ as

$$\mathbf{D}'W^*\mathbf{D} = \mathbf{D}_0'W_0^*\mathbf{D}_0 + \mathbf{D}_1'W_1^*\mathbf{D}_1 + r_n, \tag{10}$$

where $W_0^*$ is the $(n-2k)\times(n-2k)$ middle tri-diagonal sub matrix of $W^*$, $\mathbf{D}_0$ is the set of $D$'s that correspond to the sub-matrix $W_0^*$ and

$$W_1^* = \begin{pmatrix} A_{k\times k} & \mathbf{0} \\ \mathbf{0} & B_{k\times k} \end{pmatrix},$$

$\mathbf{D}_1$ is the set of $D$'s that correspond to the two sub-matrices $A$ and $B$ at top left corner and, bottom right corner of $W^*$ and $r_n$ is the sum of the terms that would be left out in this extraction (namely, the terms contributed by matrices $\mathbf{0}_i, i = 1, 2, 3, 4$).

Our approach is to show that under $H_0$, a properly normalized version of $\mathbf{D}_0'W_0^*\mathbf{D}_0$ converges in distribution to a standard normal random variable while the other terms on the right side of (10) converge in probability to 0 when divided by the normalizing quantity for $\mathbf{D}_0'W_0^*\mathbf{D}_0$. First note that,

$$\mathbf{D}_0'W_0^*\mathbf{D}_0 = -\sum_{i=k+1}^{n-k-1} D_iD_{i+1}. \tag{11}$$

Then, $E_0(D_iD_{i+1}) = 0$ and $Var_0(D_iD_{i+1}) = \sigma^4$ and, $Cov_0(D_iD_{i+1}, D_jD_{j+1}) = 0$. Therefore $Var_0(\mathbf{D}_0'W_0^*\mathbf{D}_0) = (n-k-1)\sigma^4$. Since we assume that the fourth moments of $D_i$'s are finite, under $H_0$, Central Limit Theorem can be applied to get

$$\frac{\mathbf{D}_0' W_0^* \mathbf{D}_0}{\sqrt{(n-k-1)\sigma^4}} \xrightarrow{d} N(0,1),$$

as $n \to \infty$. Now, since $\hat{\sigma^2}$ is a consistent estimator of $\sigma^2$, we can replace $\sigma$ by $\hat{\sigma}$ in the above without changing the limiting distribution. Also, since $(n-2k)/n \to 1$ as $n \to \infty$, we have

$$\frac{\mathbf{D}_0' W_0^* \mathbf{D}_0}{\sqrt{n\hat{\sigma^4}}} \xrightarrow{d} N(0,1),$$

as $n \to \infty$.

It remains to show that $\frac{\mathbf{D}_1' W_1^* \mathbf{D}_1 + r_n}{\sqrt{n\sigma^4}}$ converges in probability to 0. It is clear that $r_n/\sqrt{n}$ converges in probability to 0.

Now, $E_0(r_n) = 0$. Also, noting that, $E_0(T) = \sum_{i=1}^{n-1} E_0(\hat{\psi_0}(u_i)) = 0$, we get $E_0(\mathbf{D}_1' W_1^* \mathbf{D}_1) = 0$. Now, consider the matrix $W_1^*$. In $W_1^*$, if we replace both the $k \times (k+1)$ and $(k+1) \times k$ elements by $-1/2$, it becomes a symmetric tridiagonal matrix for which $E_0(\mathbf{D}_1' W_1^* \mathbf{D}_1) = 0$, while the asymptotic properties of the quadratic form $\frac{\mathbf{D}_1' W_1^* \mathbf{D}_1}{\sqrt{n\sigma^4}}$ do not change. Thus, we will consider this modified form as $W_1^*$ hereafter. We can diagonalize the quadratic form $\mathbf{D}_1' W_1^* \mathbf{D}_1$ as

$$\mathbf{D}_1' W_1^* \mathbf{D}_1 = \sum_{i=1}^{2k} \beta_i \chi_i^2,$$

where $\beta_i, i = 1, ..., 2k$ are eigenvalues of $W_1^*$. It is not difficult to show that $|\beta_i| \leq M$ for some $M$. Thus, if we define $V_i = \beta_i \chi_i^2$, we can apply an iterated law of logarithm to the sum $\sum_{i=1}^{2k} V_i$ to get

$$\frac{\sum_{i=1}^{2k} V_i}{\sqrt{2k[\ln(2k)]^{(\xi+1)}}} \xrightarrow{a.s.} 0$$

as $k \to \infty$ (Chung, 1974). Here $0 < \xi < 1$. Since

$$\frac{2k[\ln(2k)]^{(\xi+1)}}{n} \to 0$$

as $n \to \infty$, we have the desired result.

**Proof of Theorem 2.2**

We sketch the proof for scheme 1. The proof for second scheme is very similar.

We will consider $0 < u < 1$ and $l = 1$. The other values of $u$ and $l$ can be dealt similarly. Consider the situation where the $m$ ordered design points $u_i, i = 1, .., m$ partion the interval $[0, 1]$ into $m$ intervals each with length of order

$1/m$. Let $r = n - m$. If $r$ is very small compared with $m$ for large $m$, (following the assumptions), the number of pseudo $D$'s in any given interval $[u_i, u_{i+1}]$ would be at most 1. Then with $l = 1$, we may take pseudo $\tilde{D}_j$'s to be of the form $\tilde{D}_j = (D_i + D_{i+1})/2$.

Assuming that there is at most one pseudo observation in between two original $D$'s, we first examine the contribution of $\tilde{D}_j$'s, $j = m+1, .., n$ to the sum $-\sum_{i=1}^{n-1} D_i D_{i+1}$ using $D_i, i = 1, ..., m$ and the pseudo $\tilde{D}_j$'s $j = m+1, ..., n$. Suppose there is a $\tilde{D}$ between $u_1$ and $u_2$. Then, we would have a term $D_1\tilde{D} + \tilde{D}D_2$ in the above sum instead of just $D_1D_2$. Now, $D_1\tilde{D} + \tilde{D}D_2 = D_1D_2 + (D_1^2 + D_2^2)/2$. Hence, whenever a $\tilde{D}$ is added into an interval $[u_i, u_{i+1}]$, that would give a term of type $(D_1^2 + D_2^2)/2$ in addition to $-\sum_{i=1}^{m-1} D_i D_{i+1}$. Without loss of generality, we can assume that that all pseudo $D$'s are in consecutive intervals $[u_i, u_{i+1}]$, $i = s, ...., s+r$. Then, $-\sum_{i=1}^{n-1} D_i D_{i+1}$ can be written as $-\sum_{i=1}^{m-1} D_i D_{i+1} + A_r$ where

$$A_r = [D_s^2 + 2D_{s+1}^2 + ..... + 2D_{s+r-1}^2 + D_{s+r}^2]/2.$$

The differences $D_s, ...., D_{s+r}$ are independent of each other. Then, writing $\frac{\sum D_i^2}{\sqrt{m}}$ as $\frac{\sum D_i^2 - \sigma^2 + \sigma^2}{\sqrt{m}}$, it can be seen that $A_r/\sqrt{m}$ converges in probability to 0 as $m \to \infty$ due to the assumption that $r/\sqrt{m} \to 0$ as $m \to \infty$. Thus, one may only consider $-\sum_{i=1}^{m-1} D_i D_{i+1}$ in creating the equation (11) above. Then. noting that $n/m \to 1$ as $m \to \infty$, the proof follows along the same lines as in the previous theorem.

## REFERENCES


K. L. Chung, A Course in Probability Theory, Academic Press, New York(1974).

M.A. Delgado, Testing the Equality of Non Parametric Regression Curves, Statistics & Probability Letters, 17(1993), 199-204.

H. Dette and S. Derbort, Analysis of Variance in Nonparametric Regression, Journal of Multivaraite Analysis(2001), 110-137.

J. Fan, Design Adaptive Nonparametric Regression, Journal of the American Statistical Association, 87(1992), 998-1004.

P. Hall and J.D. Hart, Bootstrap Test for Difference Between Means in Nonparametric Regression, Journal of the American Statistical Association, 85(1990), 1039-1049.

P. Hall, J.W. Kay, and D.M. Titterington, Asymptotically Optimal Difference Based Estimation of Variance in Nonparametric Regression, Biometrika, 77(1990), 521-528.

P. Hall and B.W. Turlach, Interpolation Methods for Adapting to Sparse Design in Nonparametric Regression, Journal of the American Statistical Association, 90(1995), 1074-1083.

P. Hall, C. Huber and P.L. Speckman,Covariate-matched One-sided Tests for the Difference Between Functional Means, Journal of the American Statistical Association, 92(1997), 1074-1083.

J.D. Hart, Nonparametric Smoothing and Lack-of-Fit Tests, Springer, New York (1997).

E. King, J.D. Hart and T.E.Wehrly, Testing the Equality of Two Regression Curves Using Linear Smoothers, Statistics & Probability Letters 12(1991), 239-247.

K.B. Kulasekera, Comparison of Regression Curves Using Quasi Residuals, Journal of the American Statistical Association 90(1995), 1085-1093.

K.B. Kulasekera and C.M. Gallagher, Variance Estimation in Nonparametric Multiple Regression, Communications in Statistcs: Theory and Methods 31(2001),1373-1383.

K.B. Kulasekera and J. Wang, Smoothing Parameter Selection for Power Optimality in Testing of Regression Curves, Journal of the American Statistical Association 92(1997), 500-511.

Linton and W. Härdle (1996), Estimation of Additive Regression Models with Known Links, Biometrika, 83(1996), 529-540.

E.A. Nadaraya, On Estimating Regression, Theory Probab. Appli. 9(1964) 141-142.

P.E. Thall and S.C. Vail, Some Covariance Models for Longitudinal Count Data with Over dispersion, Biometrics 46(1990), 657-671.

G.S. Watson, Smooth Regression Analysis, *Sankya* (Series A) 26(1964), 359-372.

Table 2: Empirical Power times 1000 for 0.05 Tests with Standard Normal Errors; $n = 50$

| Functions | Bandwidth | Projection $u = x_1$ | Projection $u = x_2$ | Projection $u = x_1 + x_2$ | DT |
|---|---|---|---|---|---|
| $f(\mathbf{x}) = x_1^2 + x_2^2 + x_1 x_2, g(\mathbf{x}) = f(\mathbf{x})$ | 0.05 | 59 | 54 | 48 | 49 |
| | 0.1 | 65 | 69 | 68 | |
| | 0.2 | 60 | 59 | 67 | |
| $f(\mathbf{x}) = x_1^2 + e^{x_1 x_2}, g(\mathbf{x}) = f(\mathbf{x})$ | 0.05 | 63 | 62 | 74 | 44 |
| | 0.1 | 53 | 54 | 61 | |
| | 0.2 | 72 | 59 | 68 | |
| $f(\mathbf{x}) = \sin(2\pi x_1) + \cos(2\pi x_2)$ | 0.05 | 63 | 61 | 67 | 35 |
| $g(\mathbf{x}) = f(\mathbf{x})$ | 0.1 | 62 | 62 | 65 | |
| | 0.2 | 56 | 60 | 76 | |
| $f(\mathbf{x}) = \sin(2\pi x_1) + \cos(2\pi x_2)$ | 0.05 | 850 | 861 | 856 | 126 |
| $g(\mathbf{x}) = \cos(2\pi x_1) + \sin(2\pi x_2)$ | 0.1 | 847 | 832 | 867 | |
| | 0.2 | 842 | 859 | 868 | |
| $f(\mathbf{x}) = \sin(2\pi x_1)$ | 0.05 | 311 | 314 | 290 | 76 |
| $g(\mathbf{x}) = \sin(2\pi x_1) + \cos(2\pi x_2)$ | 0.1 | 282 | 321 | 289 | |
| | 0.2 | 283 | 270 | 274 | |
| $f(\mathbf{x}) = \sin(2\pi x_1) + \cos(2\pi x_2)$ | 0.05 | 684 | 672 | 697 | 126 |
| $g(\mathbf{x}) = \sin(2\pi x_1) \cos(2\pi x_2)$ | 0.1 | 697 | 681 | 718 | |
| | 0.2 | 698 | 700 | 732 | |
| $f(\mathbf{x}) = x_1^3 + 2x_2^3$ | 0.05 | 588 | 569 | 579 | 724 |
| $g(\mathbf{x}) = 3x_1^2 + x_2^2$ | 0.1 | 604 | 573 | 623 | |
| | 0.2 | 598 | 561 | 616 | |

Table 3: Empirical Power times 1000 for 0.05 Tests with $t_6$ Errors; $n = 50$

| Functions | Bandwidth | Projection $u = x_1$ | Projection $u = x_2$ | Projection $u = x_1 + x_2$ | DT |
|---|---|---|---|---|---|
| $f(\mathbf{x}) = x_1^2 + x_2^2 + x_1 x_2, g(\mathbf{x}) = f(\mathbf{x})$ | 0.05 | 61 | 55 | 54 | 52 |
| | 0.1 | 74 | 65 | 54 | |
| | 0.2 | 62 | 55 | 59 | |
| $f(\mathbf{x}) = x_1^2 + e^{x_1 x_2}, g(\mathbf{x}) = f(\mathbf{x})$ | 0.05 | 54 | 61 | 69 | 47 |
| | 0.1 | 59 | 56 | 70 | |
| | 0.2 | 52 | 65 | 62 | |
| $f(\mathbf{x}) = \sin(2\pi x_1) + \cos(2\pi x_2)$ | 0.05 | 62 | 63 | 65 | 39 |
| $g(\mathbf{x}) = f(\mathbf{x})$ | 0.1 | 68 | 67 | 83 | |
| | 0.2 | 59 | 61 | 81 | |
| $f(\mathbf{x}) = \sin(2\pi x_1) + \cos(2\pi x_2)$ | 0.05 | 626 | 614 | 629 | 108 |
| $g(\mathbf{x}) = \cos(2\pi x_1) + \sin(2\pi x_2)$ | 0.1 | 630 | 617 | 642 | |
| | 0.2 | 644 | 650 | 658 | |
| $f(\mathbf{x}) = \sin(2\pi x_1)$ | 0.05 | 219 | 195 | 186 | 68 |
| $g(\mathbf{x}) = \sin(2\pi x_1) + \cos(2\pi x_2)$ | 0.1 | 195 | 218 | 221 | |
| | 0.2 | 177 | 194 | 230 | |
| $f(\mathbf{x}) = \sin(2\pi x_1) + \cos(2\pi x_2)$ | 0.05 | 524 | 480 | 506 | 99 |
| $g(\mathbf{x}) = \sin(2\pi x_1)\cos(2\pi x_2)$ | 0.1 | 540 | 505 | 503 | |
| | 0.2 | 513 | 475 | 527 | |
| $f(\mathbf{x}) = x_1^3 + 2x_2^3$ | 0.05 | 360 | 352 | 381 | 541 |
| $g(\mathbf{x}) = 3x_1^2 + x_2^2$ | 0.1 | 376 | 370 | 358 | |
| | 0.2 | 346 | 345 | 365 | |

Table 4: Empirical Size times 1000 for 0.05 Tests with Standard Normal Errors; $n = 64$

| Functions | Bandwidth | Projection | | | | DT |
|---|---|---|---|---|---|---|
| | | $u = x_1$ | $u = x_2$ | $u = x_3$ | $u = x_1 + x_2 + x_3$ | |
| $f(\mathbf{x}) = x_1^2 + 3x_2 + 2\sqrt{x_3}, g(\mathbf{x}) = f(\mathbf{x})$ | 0.05 | 60 | 59 | 57 | 65 | 45 |
| | 0.1 | 59 | 52 | 63 | 69 | |
| | 0.2 | 49 | 55 | 64 | 60 | |
| $f(\mathbf{x}) = \sin(2\pi x_1 x_2 x_3) e^{x_1+x_2+x_3}, g(\mathbf{x}) = f(\mathbf{x})$ | 0.05 | 62 | 63 | 64 | 49 | 49 |
| | 0.1 | 74 | 61 | 68 | 48 | |
| | 0.2 | 58 | 57 | 51 | 69 | |
| $f(\mathbf{x}) = \sin(2\pi x_1) + \cos(2\pi x_2) + \sin(2\pi x_3), g(\mathbf{x}) = f(\mathbf{x})$ | 0.05 | 58 | 57 | 71 | 65 | 49 |
| | 0.1 | 60 | 62 | 78 | 69 | |
| | 0.2 | 62 | 59 | 61 | 81 | |

Table 5: Empirical Size times 1000 for 0.05 Tests with $t_6$ Errors; $n = 64$

| Functions | Bandwidth | Projection | | | | DT |
|---|---|---|---|---|---|---|
| | | $u = x_1$ | $u = x_2$ | $u = x_3$ | $u = x_1 + x_2 + x_3$ | |
| $f(\mathbf{x}) = x_1^2 + 3x_2 + 2\sqrt{x_3}, g(\mathbf{x}) = f(\mathbf{x})$ | 0.05 | 62 | 60 | 57 | 57 | 72 |
| | 0.1 | 68 | 63 | 65 | 56 | |
| | 0.2 | 67 | 57 | 66 | 60 | |
| $f(\mathbf{x}) = \sin(2\pi x_1 x_2 x_3) e^{x_1+x_2+x_3}, g(\mathbf{x}) = f(\mathbf{x})$ | 0.05 | 64 | 80 | 80 | 92 | 61 |
| | 0.1 | 71 | 81 | 71 | 101 | |
| | 0.2 | 61 | 71 | 78 | 109 | |
| $f(\mathbf{x}) = \sin(2\pi x_1) + \cos(2\pi x_2) + \sin(2\pi x_3), g(\mathbf{x}) = f(\mathbf{x})$ | 0.05 | 66 | 70 | 55 | 71 | 43 |
| | 0.1 | 61 | 77 | 63 | 74 | |
| | 0.2 | 54 | 72 | 59 | 76 | |

Table 6: Empirical Power times 1000 for 0.05 Tests with Standard Normal Errors; $n = 64$

| | | Projection | | | | DT |
|---|---|---|---|---|---|---|
| Functions | Bandwidth | $u = x_1$ | $u = x_2$ | $u = x_3$ | $u = x_1 + x_2 + x_3$ | |
| $f(\mathbf{x}) = x_1^2 + 3x_2 + 2\sqrt{x_3}$ | 0.05 | 276 | 287 | 257 | 286 | 148 |
| $g(\mathbf{x}) = 2x_1 + x_2 + 3x_3$ | 0.1 | 255 | 297 | 261 | 293 | |
| | 0.2 | 268 | 303 | 274 | 299 | |
| $f(\mathbf{x}) = 2 * \sin(2\pi x_1 x_2 x_3)$ | 0.05 | 906 | 912 | 914 | 914 | 285 |
| $g(\mathbf{x}) = 2 * \cos(2\pi x_1 x_2 x_3)$ | 0.1 | 903 | 912 | 916 | 922 | |
| | 0.2 | 901 | 905 | 908 | 931 | |
| $f(\mathbf{x}) = \sin(2\pi x_1) + \cos(2\pi x_2) + \sin(2\pi x_3)$ | 0.05 | 783 | 782 | 767 | 765 | 116 |
| $g(\mathbf{x}) = \cos(2\pi x_1) + \sin(2\pi x_2) + \cos(2\pi x_3)$ | 0.1 | 755 | 774 | 782 | 791 | |
| | 0.2 | 783 | 780 | 779 | 798 | |
| $f(\mathbf{x}) = 2 * \sin(2\pi(x_1 + x_2 + x_3))$ | 0.05 | 120 | 149 | 153 | 129 | 68 |
| $g(\mathbf{x}) = 2 * \cos(2\pi(x_1 + x_2 + x_3))$ | 0.1 | 120 | 155 | 158 | 140 | |
| | 0.2 | 110 | 145 | 145 | 148 | |

Table 7: Empirical Power times 1000 for 0.05 Tests with $t_6$; $n = 64$

| Functions | Bandwidth | Projection $u = x_1$ | $u = x_2$ | $u = x_3$ | $u = x_1 + x_2 + x_3$ | DT |
|---|---|---|---|---|---|---|
| $f(\mathbf{x}) = x_1^2 + 3x_2 + 2\sqrt{x_3}$ | 0.05 | 179 | 206 | 164 | 210 | 78 |
| $g(\mathbf{x}) = 2x_1 + x_2 + 3x_3$ | 0.1 | 191 | 210 | 169 | 221 | |
| | 0.2 | 180 | 191 | 169 | 235 | |
| $f(\mathbf{x}) = 2 * \sin(2\pi x_1 x_2 x_3)$ | 0.05 | 788 | 799 | 811 | 802 | 249 |
| $g(\mathbf{x}) = 2 * \cos(2\pi x_1 x_2 x_3)$ | 0.1 | 789 | 813 | 802 | 811 | |
| | 0.2 | 766 | 788 | 798 | 816 | |
| $f(\mathbf{x}) = \sin(2\pi x_1) + \cos(2\pi x_2) + \sin(2\pi x_3)$ | 0.05 | 594 | 532 | 608 | 564 | 93 |
| $g(\mathbf{x}) = \cos(2\pi x_1) + \sin(2\pi x_2) + \cos(2\pi x_3)$ | 0.1 | 614 | 552 | 612 | 581 | |
| | 0.2 | 585 | 540 | 597 | 580 | |
| $f(\mathbf{x}) = 2 * \sin(2\pi(x_1 + x_2 + x_3))$ | 0.05 | 114 | 111 | 128 | 119 | 66 |
| $g(\mathbf{x}) = 2 * \cos(2\pi(x_1 + x_2 + x_3))$ | 0.1 | 128 | 109 | 135 | 123 | |
| | 0.2 | 109 | 104 | 121 | 134 | |

Table 8: Average and Sample Variance of 1000 Estimates of the Error Variance

| Regression Function | Error Variance= 1 | | |
|---|---|---|---|
| | $n = 25$ | $n = 64$ | $n = 100$ |
| $\mu(\mathbf{t}) = t_1^2 + t_2^3 + t_1 t_2$ | 1.1016 | 1.0904 | 1.0624 |
| | 0.1378 | 0.0519 | 0.0350 |
| $\mu(\mathbf{t}) = t_1 e^{t_1 t_2}$ | 1.0474 | 1.0296 | 1.0213 |
| | 0.1320 | 0.0482 | 0.0321 |
| $\mu(\mathbf{t}) = \sin(2\pi t_1) + \cos(2\pi t_2)$ | 1.4387 | 1.2391 | 1.1798 |
| | 0.2070 | 0.0647 | 0.0361 |
| Regression Function | $n = 27$ | $n = 81$ | $n = 125$ |
| $\mu(\mathbf{t}) = t_1 + t_2 + t_3$ | 1.1250 | 1.0879 | 1.0678 |
| | 0.1431 | 0.0544 | 0.0375 |
| $\mu(\mathbf{t}) = \sin(2\pi t_1 t_2 t_3)$ | 1.4975 | 1.4766 | 1.4856 |
| | 0.2441 | 0.0971 | 0.0470 |
| $\mu(\mathbf{t}) = \sin(2\pi t_1) + \cos(2\pi t_2) + sin(2\pi t_3)$ | 1.6128 | 1.5720 | 1.4804 |
| | 0.3712 | 0.1179 | 0.0409 |