# Quantifying the cost of simultaneous non-parametric approximation of several samples

P. L. Davies*
University of Duisburg-Essen, Germany
Technische Universiteit Eindhoven, Holland

A. Kovac*
University of Bristol, United Kingdom

December 7, 2006

**Abstract**

We consider the standard non-parametric regression model with Gaussian errors but where the data consist of different samples. The question to be answered is whether the samples can be adequately represented by the same regression function. To do this we define for each sample a universal, honest and non-asymptotic confidence region for the regression function. Any subset of the samples can be represented by the same function if and only if the intersection of the corresponding confidence regions is non-empty. If the empirical supports of the samples are disjoint then the intersection of the confidence regions is always non–empty and a negative answer can only be obtained by placing shape or quantitative smoothness conditions on the joint approximation. Alternatively a simplest joint approximation function can be calculated which gives a measure of the cost of the joint approximation, for example, the number of extra peaks required.

*Key Words and Phrases.* Modality, nonparametric regression, penalization, regularization, total variation.

*Research supported in part by Sonderforschungsbereich 475, University of Dortmund

# 1 Introduction

We consider the non-parametric regression model for $k$ samples

$$Y_i(t) = f_i(t) + \sigma_i Z_i(t),\ i = 1, \ldots, k, \quad t \in [0, 1] \tag{1}$$

where the $Z_i$ are independent standard Gaussian white noise processes. The problem is that of deciding whether $k$ samples

$$\boldsymbol{y}_{in_i} = \{(t_{ij}, y_{ij}) : j = 1, \ldots, n_i\},\ i = 1, \ldots, k \tag{2}$$

with supports

$$S_i = \{t_{i1} < t_{i2} < \ldots < t_{in_i}\},\ i = 1, \ldots, k. \tag{3}$$

can be simultaneously represented by the same function $f$. In the following we restrict attention to the case $k = 2$; the extention to more samples poses no problems. The standard approach is to consider the null hypothesis

$$H_0 : f_1 = f_2 \quad H_1 : f_1 \neq f_2 \tag{4}$$

and then to construct a test which is asymptotically consistent and can detect alternatives which converge to the null hypothesis at certain rates $f_1(t) - f_2(t) = \Delta(t)/\gamma(n)$. The best result seems to be that of Neumeyer and Dette (2003) who construct a test which can detect alternatives which converge to the null hypothesis at the optimal rate $\gamma(n) = n^{1/2}$. If the supports $S_i\,, i = 1, 2$, are equal then it is not difficult to construct such a test as the differences $Y_1(t_j) - Y_2(t_j)$ do not depend on $f$ (see for example Delgado (1992) and Fan and Lin (1998)). The result of Neumeyer and Dette (2003) holds however even if the supports are disjoint, $S_1 \cap S_2 = \emptyset$. We point out that in this case there are certain difficulties which can be most clearly seen in the case of exact data

$$y_{ij} = f_i(t_{ij}),\ t_{ij} \in S_i, \quad i = 1, 2.$$

If we denote the supremum norm on $[0, 1]$ by $\| \|_\infty$ then the null and alternative hypotheses of (4) may be rewritten as

$$H_0 : \|f_1 - f_2\|_\infty = 0, \quad H_1 : \|f_1 - f_2\|_\infty > 0. \tag{5}$$

If the values of $f_1$ and $f_2$ are known only on disjoint sets $S_1$ and $S_2$ respectively then it is not possible to decide between $H_0$ and $H_1$. This continues to hold even if $f_1$ and $f_2$ are subject to qualitative smoothness conditions such as infinite differentiability. Even in this case a function can be found which simply interpolates the data. All conditions imposed in the literature are of this form: Hall and Hart (1990), a bounded first derivative; Härdle and Marron (1990), Hölder continuity; King, Hart and Wehrly (1990), 'at least uniform continuity; Kukasekera (1995), Kukasekera and Wang (1997), a continuous second derivative; Munk and Dette (1998), Hölder continuity of order $\gamma > 1/2$; Dette and Neumeyer (2001), a continuous $r$th derivative: Lavergne (2001), a second derivative which is uniformly Lipschitz of order $\gamma$, $0 \leq \gamma < 1$; Neumeyer and Dette (2003), continuous derivatives of order $d \geq 2$. In order to distinguish between $H_0$ and $H_1$ it is necessary to place either quantitative conditions on $f_1$ and $f_2$ such as first derivatives bounded by a specified value, for example $\|f_1^{(1)}\|_\infty \leq 1$, $\|f_2^{(1)}\|_\infty \leq 1$, or shape restrictions such as $f_1$ and $f_2$ being monotone.

These conclusions are in a sense trivial but as far as we are aware they have never been explicitly mentioned in the literature. The addition of noise does not alter anything. However different the functions $f_1$ and $f_2$, whatever the sample sizes and whatever the value of the test statistic there will always be a single common function which is not excluded by the smoothness conditions and which is consistent with both samples. All that can be said is that if the functions $f_1$ and $f_2$ are different then any common function $g$ equal to $f_i$ on $S_i$, $i = 1, 2$ becomes more complicated as the sample sizes increase. It is this increase in complexity which we call the cost of the simultaneous approximation. In the remainder of the paper we show how this can be quantified. Our approach can be split into two parts.

(1) Firstly for each sample $\boldsymbol{y}_{in_i} = \{(t_{ij}, y_{ij}) : j = 1, \ldots, n_i\}$ we specify a so called approximation region $\mathcal{A}_{in_i}$ which specifies those functions $f_i$ for which the model (1) is an adequate approximation for the sample. The intersection of the approximation regions $\mathcal{A}_{1,n_1} \cap \mathcal{A}_{2,n_2}$ contains all those functions which simultaneously approximate both samples. It is also the approximation region for the simultaneous approximation. A similar idea in the context of the one-way table in the analysis of variance is expounded in Davies (2004).

(2) Secondly using some measure of complexity we regularize within each

approximation region by choosing the simplest function.

The idea of approximation followed by regularization may be found in Davies and Kovac (2001, 2004) but it is made much more explicit here. In particular if the data are generated under the model (1) then we show that the approximation regions are universal, honest and non-asymptotic confidence regions. We use the word 'approximation' as we are not trying to estimate the 'true' function $f$ but are rather concerned with calculating a simplest function consistent with the data (see Donoho (1988) and Davies (1995)).

In Section 2 we define the approximation or confidence regions and consider the problem of regularization in Section 3. Finally in Section 4 we apply the ideas and concepts to the problem of comparing regression functions.

## 2 Approximation regions

We consider a single sample of data $\boldsymbol{Y}_n = (t_i, Y(t_i))_1^n$ generated under the model

$$Y(t) = f(t) + \sigma Z(t) \tag{6}$$

where we take the $t_i$ to be ordered. For any function $g$ and interval $I \subset [0, 1]$ we put

$$w(g, \boldsymbol{Y}_n, I) = \frac{1}{\sqrt{|I|}} \sum_{t_i \in I} (Y(t_i) - g(t_i)) \tag{7}$$

where $|I|$ denotes the number of points $t_i \in I$. The approximation or confidence region $\mathcal{A}_n$ is defined by

$$\mathcal{A}_n(\boldsymbol{Y}_n, \mathcal{I}_n, \sigma, \tau_n) = \{g : \max_{I \in \mathcal{I}_n} |w(g, \boldsymbol{Y}_n, I)| \ \leq \sigma \sqrt{\tau_n \log(n)} \ \}. \tag{8}$$

where $\mathcal{I}_n$ is a collection of intervals of [0, 1] and typically $\tau_n > 2$. Very often it is sufficient for $\mathcal{I}_n \cap \{t_1, \ldots, t_n\}$ to correspond to a wavelet multiresolution scheme as follows. It contains all singletons $\{t_i\}$, the pairs $\{t_1, t_2\}$, $\{t_3, t_4\}, \ldots$, quadruples $\{t_1, t_2, t_3, t_4\}, \{t_5, t_6, t_7, t_8\}$ etc. with any set left over at the end also being included. To simplify calculations however we take $\mathcal{I}_n$ to be the set of all intervals. As

$$w(f, \boldsymbol{Y}_n, I) = \frac{1}{\sqrt{|I|}} \sum_{t_i \in I} Z(t_i)$$

does not depend on $f$ we see that

$$\boldsymbol{P}(f \in \mathcal{A}_n(\boldsymbol{Y}_n, \mathcal{I}_n, \sigma, \tau_n))$$

also does not depend on $f$. For any given $\alpha$ and collections of intervals $\mathcal{I}_n$ we may choose $\tau_n$ so that

$$\boldsymbol{P}\left(\max_{I\in\mathcal{I}_n} \frac{1}{\sqrt{|I|}}\Big|\sum_{i\in I} Z(t_i)\Big| \le \sqrt{\tau_n \log n}\right) = 1-\alpha.$$

For this choice of $\tau_n$ we have

$$\boldsymbol{P}(f \in \mathcal{A}_n(\boldsymbol{Y}_n, \mathcal{I}_n, \sigma, \tau_n)) = \alpha. \tag{9}$$

so that $\mathcal{A}_n$ is an exact confidence region of size $1-\alpha$ for $f$ with no restrictions on $f$. The confidence region (8) treats all intervals equally but we may define a second one which downweights the importance of small intervals as follows. Dümbgen and Spokoiny (2001) have extended Lèvy's uniform modulus of continuity of the Brownian motion and shown that

$$\sup_{0<s<t<1} \frac{\frac{(B(t)-B(s))^2}{t-s} - 2\log(1/(t-s))}{\log(\log(e^e/(t-s)))} < \infty \quad \text{a.s.} \tag{10}$$

If we embed the partial sums $\sum_{i\in I}^{j} Z(t_i)/\sqrt{|I|}, I \in \mathcal{I}_n$ in a standard Brownian motion it follows that

$$\sup_{I\imath\mathcal{I}_n} \frac{(\sum_{t_j\in I} Z(t_j))^2/|I| - 2\log(n/|I|)}{\log(\log(e^e n/|I|)))} = \Gamma < \infty \quad \text{a.s..} \tag{11}$$

This implies that for any $\alpha$ we can find a $\gamma_n = \gamma_n(\alpha)$ such that

$$\begin{aligned}\mathcal{A}_n^{BM}(\boldsymbol{Y}_n, \mathcal{I}_n, \sigma, \gamma_n) = \{g :\ & |w(g, \boldsymbol{Y}_n, I)| \\ \le\ & \sigma\sqrt{2\log(n/|I|) + \gamma_n \log(\log(e^e n/|I|))} \text{ for all } I \in \mathcal{I}_n)\,\}.\end{aligned} \tag{12}$$

is a $(1-\alpha)$ confidence region as for $\mathcal{A}_n$.. The values of $\gamma_n$ may be determined by simulation. The index $BM$ stands for the Brownian modulus of continuity from which the confidence region derives.

The confidence regions $\mathcal{A}_n(\boldsymbol{Y}_n, \mathcal{I}_n, \sigma, \tau_n)$ and $\mathcal{A}_n^{BM}(\boldsymbol{Y}_n, \mathcal{I}_n, \sigma, \gamma_n)$ both re-

quire the true value of $\sigma$. This may be estimated from the data by putting

$$\sigma_n = \frac{1.4826}{\sqrt{2}}\text{median}\left(|Y(t_2) - Y(t_1)|, \ldots, |Y(t_n) - Y(t_{n-1})|\right) \tag{13}$$

which will be our default value in this paper. If $f$ is constant then $\sigma_n$ is a consistent estimator of $\sigma$. If $f$ is not constant then $\sigma_n$ will have a positive bias whatever $f$ and the effect of this is to make the confidence region and the inclusion probability larger. More precisely if we put

$$\tilde{\sigma}_n = \frac{1.4826}{\sqrt{2}}\text{median}(|Z(t_2) - Z(t_1)|, \ldots, |Z(t_n) - Z(t_{n-1})|) \tag{14}$$

and then choose $\tau_n$ so that

$$\boldsymbol{P}\left(\max_{I\in\mathcal{I}_n}\frac{1}{\sqrt{|I|}}\Big|\sum_{t_i\in I} Z(t_i)\Big| \le \tilde{\sigma}_n\sqrt{\tau_n\log(n)}\right) = 1-\alpha \tag{15}$$

then

$$\boldsymbol{P}(f \in \mathcal{A}_n(\boldsymbol{Y}_n, \mathcal{I}_n, \sigma_n, \tau_n)) \ge 1-\alpha. \tag{16}$$

with $\sigma_n$ given by (13). There is a corresponding result for $\mathcal{A}_n^{BM}(\boldsymbol{Y}_n, \mathcal{I}_n, \sigma_n, \gamma_n)$. The confidence region (16) is again universal and non-asymptotic but is now honest rather than exact as $1-\alpha$ is now a lower bound for the inclusion probability. In practice we simply choose a default value of $\tau_n$, for example $\tau_n = 2.3$.

It may be objected that the confidence region $\mathcal{A}_n(\boldsymbol{Y}_n, \mathcal{I}_n, \sigma, \tau_n)$ is so large that any reasonable estimate of $f$ belongs to it. This is not the case. Because the intervals $\mathcal{I}_n$ form a multiresolution scheme as described above any function $g \in \mathcal{A}_n(\boldsymbol{Y}_n, \mathcal{I}_n, \sigma, \tau_n)$ is forced to adapt at all levels to the data. Many standard methods of estimating the function $f$ are in fact not sufficiently adaptive to produce estimates which lie in $\mathcal{A}_n(\boldsymbol{Y}_n, \mathcal{I}_n, \sigma, \tau_n)$. Suppose we take $\mathcal{I}_n$ to be the set of all intervals and put $\alpha = 0.01$. Simulations give $\tau_n = 3.62$. for $n = 512$ and $\tau_n = 3.45$ for $n = 1024$. Table 1 gives the results of a small simulation study. The methods used were the taut string (TS), wavelets, a kernel estimator with a global bandwidth determined by cross validation and finally a smoothing spline with penalty parameter determined by cross validation. The reason that the taut string did not always produce an estimate in $\mathcal{A}_n(\boldsymbol{Y}_n, \mathcal{I}_n, \sigma, \tau_n)$ is that a version was used which only tested the condition

*Table 1 about here*

$$\max_{I\in\mathcal{I}_n^*}|w(g, \boldsymbol{Y}_n, I)| \ \ge \sigma_n\sqrt{2.3\log(n)}$$

on the dyadic family of intervals $\mathcal{I}_n^*$ described above. This does not check all intervals and consequently the bound $0.4\sqrt{3.45 \log 1024} = 1.96$ can be exceeded.

# 3 Regularization

The confidence region $\mathcal{A}_n(\boldsymbol{Y}_n, \mathcal{I}_n, \sigma_n, \tau_n)$ includes many functions $g$ which would not be regarded as an adequate representation of the data. In particular any function which interpolates the data belongs to the confidence region as all residuals are zero. Of interest are not such functions but those which minimize some measure of complexity $C$ subject to the function being adequate:

$$\text{minimize} \quad C(g) \quad \text{subject to} \quad g \in \mathcal{A}_n(\boldsymbol{Y}_n, \mathcal{I}_n, \sigma_n, \tau_n). \tag{17}$$

Examples are

$$C(g) = \text{ number of local extreme values} \tag{18}$$

$$C(g) = \int_0^1 g^{(2)}(t)^2 \, dt. \tag{19}$$

and

$$C(g) = TV(g^{(k)}) \tag{20}$$

where $TV$ denotes total variation. The example (18) was considered by Davies and Kovac (2001) who proposed the taut string algorithm to solve the regularization problem. The example (19) leads to a quadratic programming problem which is often numerically unstable. Davies and Meise (2005) proposed a weighted smoothing spline technique to get approximate solutions. Finally (20) leads to a linear programming problem. In general shape constraints such as (18) are much easier to interpret than smoothness constraints such as (19) and (20) and so in the remainder of the paper we restrict attention to (18).

# 4 Comparing regression functions

## 4.1 A one-way table for regression functions

We turn to the problem of comparing regression functions for samples $\boldsymbol{Y}_{in_i} = (t_{ij}, Y_i(t_{ij}))_{j=1}^{n_i}$ generated under (1). As a first step we replace the $1-\alpha$ in (15)

by $(1-\alpha)^{1/k}$ where $k$ is the number of samples. This adjusts the size of each confidence region to take into account the number of samples. The confidence region for the $i$th sample is given by

$$\mathcal{A}_{in_i} = \mathcal{A}_{in_i}(\boldsymbol{Y}_{in_i}, \mathcal{I}_{in_i}, \sigma_{in_i}, \tau_{in_i}) = \qquad (21)$$
$$\{g : \max_{I\in\mathcal{I}_{in_i}} |w(g, \boldsymbol{Y}_{in_i}, I)| \ \leq \sigma_{in_i}\sqrt{\tau_{in_i}\log(n_i)}\ \}.$$

All questions concerning the relationships between the functions $f_i$ can now be answered by using the confidence regions. For example the question as to whether the $f_i$ are all equal, the null hypothesis $H_0$ of (4) translates into the question as to whether the intersection $\mathcal{A}^k = \cap_{i=1}^k \mathcal{A}_{in_i}$ is empty or not. If the supports $S_i$ of the samples are pairwise disjoint then $\mathcal{A}_n$ is not empty and so there always is a joint approximation function. In order to obtain a negative answer the questions posed must be specific in terms of shape or, if formulated in terms of smoothness, quantitative. A quantitative smoothness constraint that the first derivatives be bounded by 2.35. These considerations are consistent with the arguments in Section 1. As argued there, asymptotics cannot overcome the lack of precision of the questions.

*Figure 1 about here*

An alternative to asking precise questions is to determine a simplest function which lies in the intersection of the approximation regions. This can be seen in Figure 1 where the measure of complexity is the number of local extreme values. The data sets were generated under the model (1) with the same sample sizes $n_1 = n_2 = 400$ and with $f_1(t) = \exp(1.5t),\ f_2(t) = \exp(1.5t) + 2$. The standard deviation of the noise was put to 0.25 in both samples and the support points $t_{ij}$ were taken to be uniformly and independently distributed on $[0, 1]$. The upper panel shows the two data sets, the centre panel shows the individual approximating functions which are both non-decreasing. The lower panel shows the joint approximating function with 347 local extreme values. The cost of the joint approximation is thus an additional 347 peaks. One can ask how close the two data sets must be before the joint approximating function is also non-decreasing. This is the case if we reduce the values of $Y_2$ by 1.8841 so that $f_1$ and $f_2$ are 0.1159 apart. Figure 2 shows the results in this case.

*Figure 2 about here*

If the supports $S_i$ of the samples are not disjoint then it may happen that the linear inequalities which define the confidence regions are inconsistent. In this case $\mathcal{A}^k = \cap_{i=1}^k \mathcal{A}_{in_i} = \emptyset$ and there is no joint approximating function. As there are many such linear inequalities this can lead to numerical problems. We

therefore adopt the following simplified procedure. For a point $t \in \cup_{i=1}^{k} S_i$ we put

$$\tilde{f}(t) = \frac{\sum_{j=1}^{k_i} Y_{i_j}(t)/\sigma^2_{i_j n_{i_j}}}{\sum_{j=1}^{k_i} 1/\sigma^2_{i_j n_{i_j}}}, \quad t \in \cap_{j=1}^{k_i} S_{i_j}, \tag{22}$$

and we check whether $\tilde{f} \in \mathcal{A}^k = \cap_{i=1}^{k} \mathcal{A}_{in_i}$. We illustrate this using the data of Figure 1. We keep the same values for the noise and for the support $S_1$ of $\boldsymbol{Y}_{1n_1}$ but set $S_2 = S_1$. It turns out that there now exists a joint approximating function for a difference $f_2 - f_1 = 0.0915$ but not for $f_2 - f_1 = 0.0916$. If one took the differences of the two samples and tested for a zero mean a difference of 0.071 could be detected. The performance of our method is in fact better than this would suggest. If one tests only for a zero mean then any function whose $g$ with $\int_0^1 g(t)\,dt$ could be added to $f_1$ to give $f_2 = f_1 + g$ and this would not be detected by differences in the means. It would however be detected by our method for a sufficiently large $g$ as we use a multiresolution scheme to analyse the residuals. The size of $g$ required is the topic of the next section.

## 4.2 Analysing the procedure

As mentioned in Section 1 the Neumeyer and Dette (2003) procedure can detect differences of the order of $n^{-1/2}$. We now consider the size of detectable differences for our procedure in the case of equal supports. For simplicity we consider only the case $k = 2$ and assume that the supports $S_1$ and $S_2$ are given by $t_{1i} = t_{2i} = i/n$. We take $\mathcal{I}_n$ to be the set of all subsets and suppose that $\sigma_1 = \sigma_2 = \sigma$. If a joint approximating function $\tilde{f}_n$ exists then for any interval $I$ of [0, 1] we have

$$\frac{1}{\sqrt{|I|}} \left| \sum_{t_i \in I} (Y_j(t_i) - f_n(t_i)) \right| \leq \sigma_{in} \sqrt{\tau_n \log(n)}, \ j = 1, 2.$$

If the estimates $\sigma_{in}$ are close to the $\sigma$ then

$$\frac{1}{\sqrt{|I|}} \left| \sum_{t_i \in I} (Y_1(t_i) - Y_2(t_i)) \right| \leq 2\sigma \sqrt{\tau_n \log(n)}.$$

For the noise we have

$$\frac{1}{\sqrt{|I|}} \left| \sum_{t_i \in I} (Z_1(t_i) - Z_2(t_i)) \right| \leq \sqrt{2}\sigma \sqrt{\tau_n \log(n)}$$

and hence

$$\frac{1}{\sqrt{|I|}}\left|\sum_{t_i\in I}(f_1(t_i)-f_2(t_i))\right| \leq 3.414\sigma\sqrt{\tau_n\log(n)}.$$

Suppose now that $f_1$ and $f_2$ differ by an amount $\gamma_n$ on an interval $I \subset [0,\,1]$, that is $f_1(t) - f_2(t) > \gamma_n, t \in I$ and that the length of $I$ is $\delta_n$. As $I$ contains about $n\delta_n$ support points points we see that

$$\frac{1}{\sqrt{n\delta_n}}n\delta_n\gamma_n \leq 3.414\sigma\sqrt{\tau_n\log(n)}$$

which implies that no joint approximation will exist if

$$\sqrt{\delta_n}\,\gamma_n > 3.414\sigma\sqrt{\tau_n\log(n)/n}. \tag{23}$$

We consider the data of Figure 1 and use the same errors and the same support $S_1$ for the first data set. We put $S_2 = S_1$ and $f_2(t) = f_1(t)$ except for the observations $201,\ldots,210$ where we put $f_2(t) = f_1(t) + \gamma_n$. For this interval $\delta_n = 10/400$ and so with $\tau_n = 2.5$ so we expect to be able to detect deviations $\gamma_n$ of the order

$$\gamma_n = 3.414\cdot 0.25\cdot\sqrt{2.5\log 400}\,/\sqrt{10} = 1.045. \tag{24}$$

In fact our method detects for this particular data a difference of 0.5812 and fails to detect a difference of 0.5811. If we put $\delta_n = 1$ in (23) so that the two functions deviate over the whole interval then

$$\gamma_n > 3.414\sigma\sqrt{\tau_n\log(n)/n}\,. \tag{25}$$

so that deviations of order $\sqrt{\log(n)/n}$ can be detected. We compare this and (23) with other procedures in the next section.

## 4.3 Comparison with other procedures

As the approach developed in this paper differs from others in the case where the supports are disjoint we restrict attention in this section to the case of equal supports. For simplicity we take $k = 2$. For such data Delgado (1992) proposed

the test statistic

$$T_n = \sqrt{n} \max_{1\le j\le n} |R(j)|/s_n^* = \max_{1\le j\le n} \left| \sum_{i=1}^{j} (Y_1(t_i) - Y_2(t_i)) \right| /(\sigma_n \sqrt{n}) \qquad (26)$$

where $\sigma_n$ is some quantifier of the noise. Under the null hypothesis $f_1 = f_2 = f$ the distribution of $T_n$ does not depend on $f$. In this special case the test statistic of Neumeyer and Dette also reduces to (26). If the data were generated under (1) with then under $H_0$ the distribution of $T_n$ converges weakly to that of $\max_{0\le t\le 1} |B(t)|$ where $B$ is a standard Brownian motion. The 0.95-quantile is approximately 2.24 which leads to rejection of $H_0$ if

$$T_n \ge 2.24. \qquad (27)$$

Suppose now that the data are generated as in (1) with $f_1(t) = f_2(t)$ apart from $t$ in an interval $I$ of length $\delta_n$ where $f_1(t) - f_2(t) \ge \gamma_n$. It follows from (27) that $H_0$ will be rejected with high probability if

$$\delta_n \gamma_n \ge 4.48\sigma/\sqrt{n} \qquad (28)$$

where $\sigma^2 = \sigma_1^2 + \sigma_2^2$. If $\delta_n = 1$ deviations of the order of $\sigma/\sqrt{n}$ can be picked up which contrasts with the $O(\sigma\sqrt{\log(n)/n}$ of (25). The difference is explained by our use of the maximal difference over all intervals whereas the statistic $T_n$ of (26) considers only intervals of the form $I = [0,\, t]$. If however $\delta_n = 1/\sqrt{n}$ it follows from (28) that the test statistic $T_n$ will pick up deviations of the order of $\sigma$ whereas it follows from (25) that out method will pick up deviations of the order of $\sigma\sqrt{\log(n)/\sqrt{n}}$. Figure 3 shows the result of applying the test (26) to the data of Figure 1 with $f_1 = f_2$ apart from the interval $[0.5025, 0.525]$ where $f_2(t) = f_1(t) + \gamma_n$. The test based on (26) requires a difference of $\gamma_n > 1.56$ if it is to be detected. As mentioned above the joint approximation detects a difference of $\gamma_n = 0.5812$.

*Figure 3 about here*

Another test which is applicable in this situation is due to Fan and Lin (1998). If we denote the Fourier transform of the data sets by $\tilde{Y}_1(i)$ and $\tilde{Y}_2(i), i = 1, \ldots, n$ ordered as described in Fan and Lin (1998) their test statistic reduces to

$$T_n^* = \sum_{1\le m\le n} \frac{1}{\sqrt{m}} \sum_{i=1}^{m} ((\tilde{Y}_2(i) - \tilde{Y}_1(i))^2/\sigma_n^2 - 1) \qquad (29)$$

where $\sigma_n$ is some estimate of the standard deviation of the $\tilde{Y}_2(i) - \tilde{Y}_1(i)$. For

data generated under the model (1) the critical value of $T_n^*$ can be obtained by simulations. It is not as simple to determine the size of the deviations which can be detected by the test (29) as the test statistic is a function of the Fourier transforms and the differences in the functions must be translated into differences in the Fourier transforms. We do not pursue this.

*Table 2 and Table 3 about here*

Tables 2 and 3 show the results of a small simulation study. We generated samples

$$Y_1(i/n) = Z_1(i/n), \quad i = 1, \ldots, n = 1000 \tag{30}$$

$$Y_2(i/n) = \gamma(i/n) + Z_2(i/n), \quad i = 1, \ldots, n = 1000 \tag{31}$$

where the $Z_j(i/n)$ are i.i.d $N(0,1)$ random variables. In Table 2 we put $\gamma(i/n) \equiv \gamma$ for various values of $\gamma$. The first line shows the percentage of cases in which a joint approximation was found using the confidence region (8). The other lines show the corresponding results for the confidence region (12), Delgado–Neumeyer–Dette test statistic (26) and the the Fan-Lin statistic (29) respectively. The results are consistent with the quantitative considerations in the last section. This form of deviation is most easily detected by the test statistics (26) and (29) (the first member of the sum in (29) is the difference of the means and this is given the largest weight) whereas the confidence regions (8) and (12) take into account differences over all the intervals $I$. As was to be expected the results for the confidence region (12) are better than those for (8) as the deviation is over the whole interval which is given more weight under (12) than under (8). Table 3 shows the results with but now with $\gamma(i/n) = \gamma$ only for the observations $i = 201, \ldots, 210$. For deviations of this kind it is to be expected that the decision based on the confidence region (8) will perform best followed by (12). This is confirmed by the results.

*Figure 4 about here*

## 4.4 An application

We give an example with some real data from the area of thin-film physics. They were kindly supplied by Professor Dieter Mergel of the University of Duisburg-Essen. Each data set is composed of 4806 measurements and the design points are the same. The samples differ in the manner in which the thin film was prepared and one of the questions to be answered is whether the results of the two methods are substantially different. The data give the intensity of reflected X-rays as a function of the angle of deflection. They are shown in

Figure 4 together with the differences $y_1(t_i) - y_2(t_i)$. The noise level for both data sets is quantified as 8.317 which is due to the fact that the data are counts of photons and hence integer valued. The differences between the two data sets are concentrated on intervals each containing about 40 observations. The estimate (28) suggests that the differences will have to be of the order of 92 to be detected with a degree of certainty by the Delgado–Neumeyer–Dette test. The actual differences are of about this order and in fact the test fails to reject the null hypothesis at the 0.1 level. The realized value of the test statistic is 1.734 as against the critical value of 1.90 given in (27). The cumulative sums of (26) are shown in Figure 5. The Fan-Lin test (29) rejects the null hypothesis at the 0.01 level. The realized value of the test statistic is 111.7 as against the critical value of 12.44 for a test of size $\alpha = 0.01$. Finally the BM test also rejects the null hypothesis at the 0.01 level. The realized value of $\Gamma$ of (11) is 53.27 as against the critical value of 0.287.

## References

Davies, P. L. and Meise, M. (2005). Approximating data with weighted smoothing splines. Technical Report, University of Duisburg-Essen, Germany.

Davies, P. L. (1995). Data features. *Statistica Neerlandica*, 49:185–245.

Davies, P. L. (2004). The one-way table: In honour of John Tukey 1915-2000. *J. Statist. Plann. Inference*, 122:3–13.

Davies, P. L. and Kovac, A. (2001). Local extremes, runs, strings and multiresolution (with discussion). *Ann. Statist.*, 29(1):1–65.

Davies, P. L. and Kovac, A. (2004). Densities, spectral densities and modality. *Ann. Statist.*, 32(3):1093–1136.

Delgado, M. A. (1992). Testing the equality of nonparametric regression curves. *Statist. Probab. Lett.*, 17:199–204.

Dette, H. and Neumeyer, N. (2001). Nonparametric analysis of covariance. *Ann. Statist.*, 29:1361–1400.

Donoho, D. L. (1988). One-sided inference about functionals of a density. *Ann. Statist.*, 16:1390–1420.

Dümbgen, L. and Spokoiny, V. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.*, 29(1):124–152.

Fan, J. and Lin, S. K. (1998). Test of significance when data are curves. *J. Amer. Statist. Assoc.*, 93:1007–1021.

Hall, P. and Hart, D. H. (1990). Bootstrap test for difference between means in nonparametric regression. *J. Amer. Statist. Assoc.*, 85(412):1039–1049.

Härdle, W. and Marron, J. S. (1990). Semiparametric comparison of regression curves. *Ann. Statist.*, 18(1):63–89.

King, E., Hart, J. D., and Wehrly, T. E. (1990). Testing the equality of two regression curves using linear smoothers. *Statist. Probab. Lett.*, 12:239–247.

Kulasekera, K. B. (1995). Comparison of regression curves using quasi-residuals. *J. Amer. Statist. Assoc.*, 90(431):1085–1093.

Kulasekera, K. B. and Wang, J. (1997). Smoothing parameter selection for power optimality in testing of regression curves. *J. Amer. Statist. Assoc.*, 92(438):500–511.

Lavergne, P. (2001). An equality test across nonparametric regressions. *J. Econometrics*, 103:307–344.

Munk, A. and Dette, H. (1998). Nonparametric comparison of several regression functions: Exact and asymptotic theory. *Ann. Statist.*, 26(6):2339–2368.

Neumeyer, N. and Dette, H. (2003). Nonparametric comparison of regression curves - an empirical process approach. *Ann. Statist.*, 31:880–920.

Arne Kovac, University of Bristol, Department of Mathematics, University Walk, Bristol, BS8 1TH, UK.

E-mail: A.Kovac@bristol.ac.uk

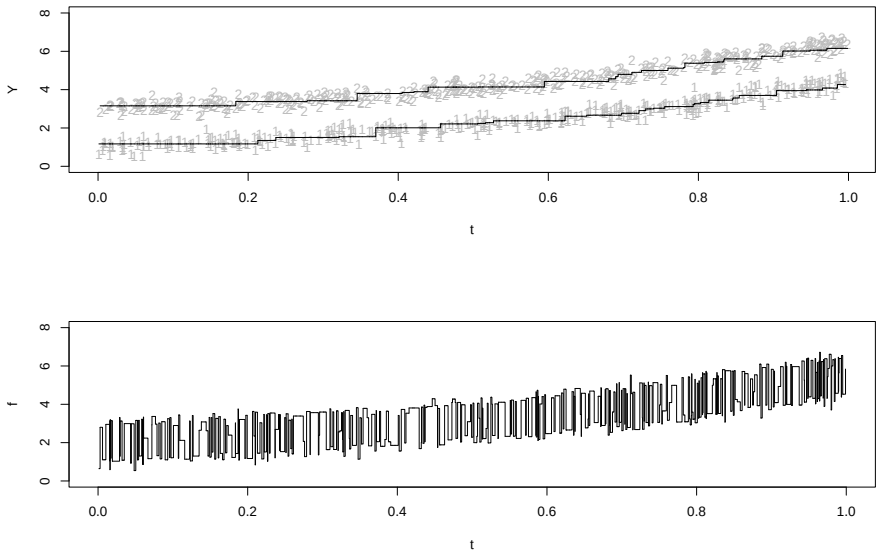


Figure 1: The upper panel shows two different samples generated by $y_1(t) = \exp(1.5t) + 0.25Z(t)$ and $Y_2(t) = \exp(1.5t) + 2 + 0.25Z(t)$ together with the approximating monotonic curves. The lower panel shows the joint approximating function with 347 local extreme values.

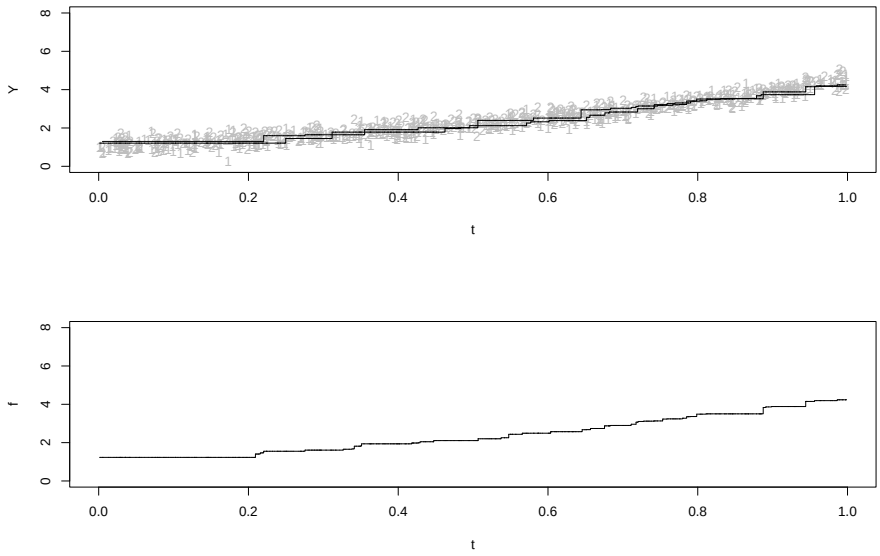


Figure 2: The upper panel shows the data of Figure 1 but with the values of $Y_2$ now given by $Y_2(t) = \exp(1.5t) + 0.1159 + 0.25Z(t)$. There is now a joint monotonic approximating function which is shown in the lower panel.

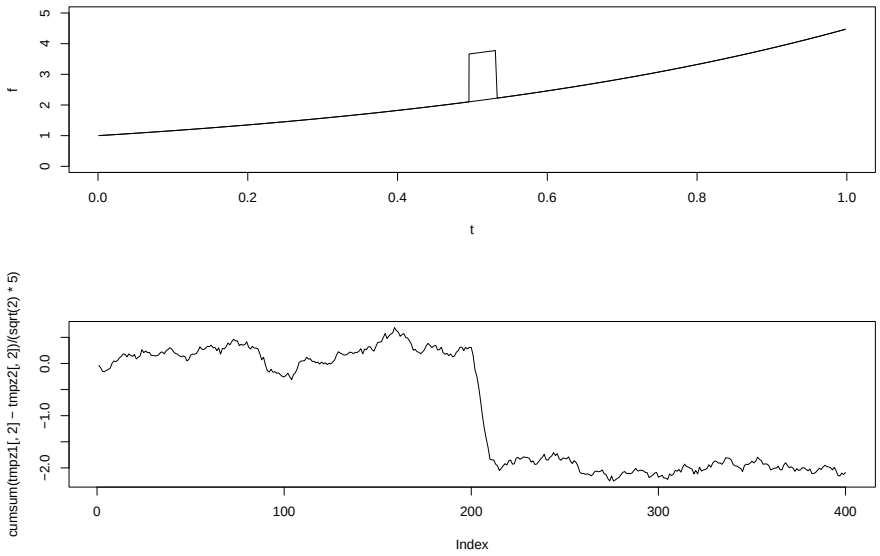


Figure 3: The upper panel shows the data of Figure 1 but with the support $T_2$ of $Y_2$ set equal to $T_1$ and with $f_2(t) = f_1(t) = \exp(1.5t)$ except on the interval $[0.5025, 0.525]$ where $f_2(t) = \exp(1.5t) + 1.56$. The cumulative sums of (26) are shown in the lower panel. The test (26) just fails to detect the difference.

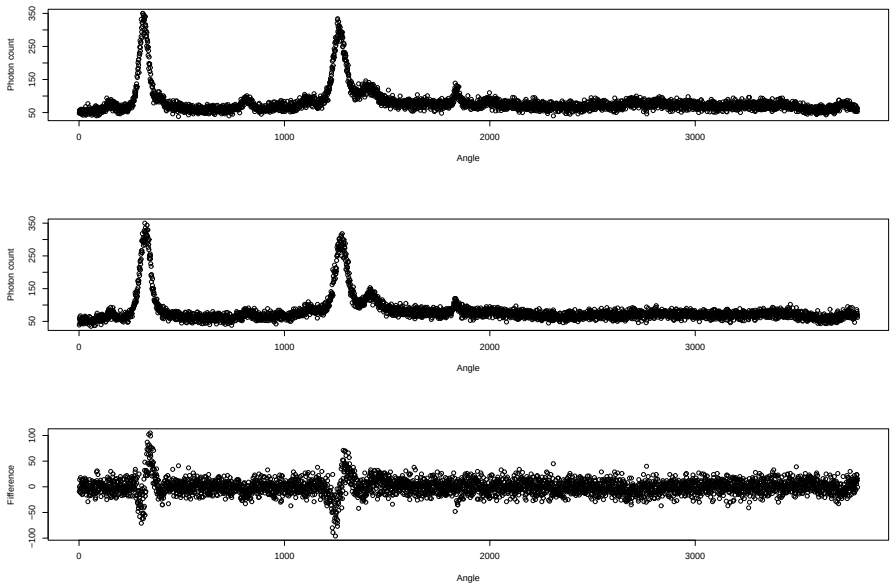


Figure 4: The top and center panels show two data sets each of 4806 observations with the same design points. The lower panel shows the differences of the two samples.

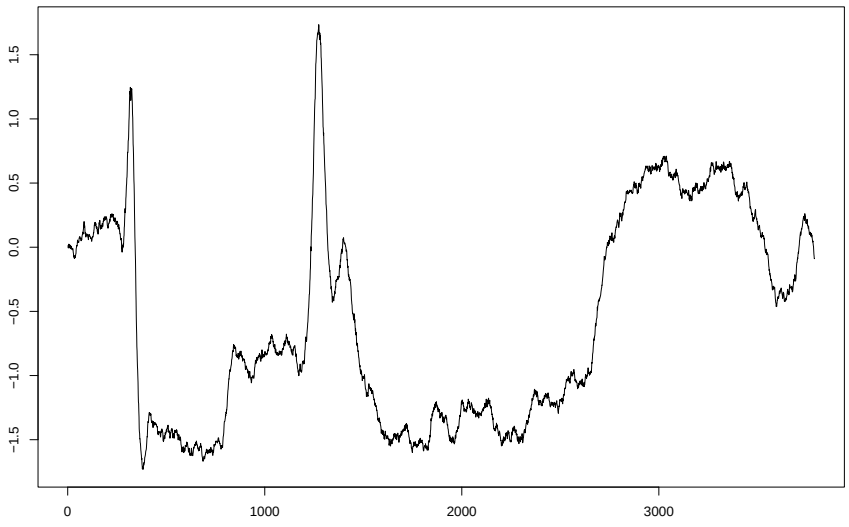


Figure 5: The cumulative sums for the test statistic (26) calculated for the thin–film data of Figure 4. The test just fails to detect any difference in the two samples. The maximum absolute value is 1.734 against a critical values of 1.90 for a test of size 0.1.

| $n = 1024$ | Bumps | Blocks | Heavisine | Doppler | Sine |
|---|---|---|---|---|---|
| TS | 96 | 98 | 100 | 88 | 100 |
| Wavelets | 0 | 0 | 0 | 0 | 100 |
| Kernel CV | 0 | 7 | 92 | 100 | 100 |
| Splines CV | 0 | 4 | 87 | 99 | 100 |

Table 1: The percentage of simulations for which the estimate of the function belonged to the approximation region $\mathcal{A}_n(\boldsymbol{Y}_n, \mathcal{I}_n, \sigma, \tau_n)$ with $n = 1024, \mathcal{I}_n$ being the set of all intervals, $\sigma = 0.4$ and $\tau_n = 3.45$.

| $\gamma$ | 0.00 | 0.05 | 0.10 | 0.15 | 0.20 | 0.25 |
|---|---|---|---|---|---|---|
| $\mathcal{A}_n$ | 4.7 | 5.6 | 5.8 | 8.4 | 11.1 | 23.8 |
| $\mathcal{A}_n^{BM}$ | 4.3 | 7.2 | 19.6 | 36.3 | 62.5 | 82.0 |
| DND | 5.0 | 9.10 | 29.2 | 52.4 | 80.2 | 94.9 |
| FL | 4.4 | 9.80 | 29.3 | 54.1 | 79.1 | 94.0 |

Table 2: The percentage of cases in which the null hypothesis $f_1 = f_2$ was rejected based on 1000 simulations for data generated according to (30) and (31) with $\gamma(i/n) = \gamma$ for all $i, 1 \leq i \leq n = 1000$. The first line refers to the decision based on the confidence region (8), the second lines that based on the confidence region (12), the third line shows the results of Delgado–Neumeyer–Dette test based on (26) and the fourth line shows the results for the Fan-Lin test based on (29).

| $\gamma$ | 0.5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 |
|---|---|---|---|---|---|---|
| $\mathcal{A}_n$ | 5.4 | 7.2 | 21.4 | 61.4 | 89.4 | 98.6 |
| $\mathcal{A}_n^{BM}$ | 5.7 | 8.5 | 20.8 | 52.6 | 82.7 | 96.7 |
| DND | 4.6 | 7.0 | 9.9 | 15.3 | 22.8 | 41.9 |
| FL | 4.6 | 6.7 | 9.8 | 13.5 | 15.8 | 23.1 |

Table 3: The percentage of cases in which the null hypothesis $f_1 = f_2$ was rejected based on 1000 simulations for data generated according to (30) and (31) with $\gamma(i/n) = \gamma$ for $i = 201 : 210$ and with $n = 1000$. The first line refers to the decision based on the confidence region (8), the second line that based on the confidence region (12), the third line shows the results of Delgado–Neumeyer–Dette test based on (26) and the fourth line shows the results for the Fan-Lin test based on (29).