Sankhyā : The Indian Journal of Statistics
Special Issue on Quantile Regression and Related Methods
2005, Volume 67, Part 2, pp 305-334

Estimation and Comparison of Fractile Graphs Using Kernel Smoothing Techniques

Bodhisattva Sen
University of Michigan, USA

Abstract

The concept of Fractile Graphical Analysis (FGA) was introduced by Prasanta Chandra Mahalanobis (see Mahalanobis, 1960). It is one of the earliest nonparametric regression techniques to compare two regression functions for two bivariate populations (X, Y). This method is particularly useful for comparing two regression functions where the covariate (X) for the two populations are not necessarily on comparable scales. For instance, in econometric studies, the prices of commodities and people's incomes observed at different time points may not be on comparable scales due to inflation. In this paper, we consider a smooth estimate of the fractile regression function and study its statistical properties. We prove the consistency and asymptotic normality of the estimated fractile regression function defined through general weight functions. We also investigate some procedures based on the idea of resampling to test the equality of the fractile regression functions for two different populations. These methods are applied to some real data sets obtained from the Reserve Bank of India, and this leads to some interesting and useful observations. In course of our investigation, we review many of Mahalanobis' original ideas relating to FGA *vis a vis* some of the key ideas used in nonparametric kernel regression.

AMS (2000) *subject classification.* Primary 62G08; secondary 62G09, 62G10.
Keywords and phrases. Fractile graphical analysis, fractile graphs, smooth estimates, weight functions, consistency and asymptotic normality, resampling methods.

1 Introduction: Fractile Graphs

Prasanta Chandra Mahalanobis introduced Fractile Graphical Analysis (FGA) as a method to compare economic data related to different populations in India over time as well as to populations differing in respect of geographical regions or in other ways. Mahalanobis used this method to study the economic condition of rural India on the basis of data collected on household consumption and expenditure over two different time periods: the 7th (October 1953 to March 1954) and the 9th (May to November 1955) rounds of the National Sample Survey of India. It is obviously of great importance to policy makers of a country like India to understand the economic condition of the rural community. They would also like to ascertain whether their policies have been able to improve the economic condition of the rural population over a period of time. As a measure of the economic well-being of the rural community, one may consider the proportion of expenditure on food articles to the total expenditure incurred. It is expected that lower this proportion, the greater is the possibility of the rural community being better off.

Let X be the total expenditure per capita per 30 days in a household and Y be the proportion of total expenditure on food articles per capita per 30 days in the household. Mahalanobis wanted to perform a regression analysis of Y on X and was interested in comparing the regression functions at two different time points. But due to inflation, the total expenditure (per capita per 30 days) for the two time points become incompatible and may cease to be comparable. Just comparing the regression functions for the two populations did not make much sense. So, he chose to compare the means of the Y-variable in different fractile groups corresponding to the X-variable. This approach leads to a novel way of standardizing the covariate X so that comparison of the two regression functions over two different time periods can be done in a more meaningful way. More precisely, FGA does this required standardization by considering $F(X)$ instead of X as the regressor, where F is the distribution function of X.

While comparing two regression functions, it is sometimes more important to understand the behaviour of the functions over a fractile interval of X and not on the entire range of X, e.g., in the example cited at the beginning, we would be more concerned with the economic condition of the bottom 5% or 10% of the population. Such localized comparison of the regression functions can also be done using FGA by restricting our attention only to the corresponding fractile intervals.

Consider two bivariate populations (X_1, Y_1) and (X_2, Y_2). We would like to compare the usual regression functions g_1 and g_2 where

$$g_1(x) = E(Y_1/X_1 = x) \qquad \text{and} \qquad g_2(x) = E(Y_2/X_2 = x).$$

When X_1 and X_2 are not in comparable scales or have their centers of distribution at different locations (for instance total household expenditure at different time points may differ significantly due to inflation), we would consider the following functions (called *fractile graph* or *fractile regression*) to draw inference about the two populations:

$$m_1(t) = E(Y_1/F_1(X_1) = t) \qquad \text{and} \qquad m_2(t) = E(Y_2/F_2(X_2) = t),$$

where F_1 and F_2 are the distribution functions of X_1 and X_2 respectively. We compare the two populations based on the behavior of the functions m_1 and m_2. FGA, therefore, differs from usual regression methods. In FGA, one is interested in the dependence of the response variable (Y_1) on the fractiles (the quantiles) of the covariate (X_1) rather than on the covariate itself.

Suppose that X is a random variable with distribution function F. Let $Z = \phi(X)$, where ϕ is any strictly increasing function on $\mathbb{R}$. Let Z have distribution function G. Then, $F(x) = G(\phi(x)) \ \forall \ x \in \mathbb{R}$, and we have $E(Y|F(X) = t) = E(Y|G(Z) = t)$. This shows that the fractile graphs remain invariant under any strictly increasing transformation of the covariate.

In Section 2 we introduce smooth estimates of fractile regression. In Section 3 we prove consistency and asymptotic normality of the estimated fractile graphs in a very general setup. In Section 4 we propose statistical methods based on resampling ideas for comparing two fractile graphs. We also discuss the multi-scale approach to curve estimation. In Section 5 we discuss the performance of the proposed testing procedures on real data as well as on simulated data. The proofs of the main results are provided in Section 6.

2 Smooth Estimation of Fractile Regression

Mahalanobis' original idea for estimating fractile graphs was as follows. Consider any bivariate population (X, Y). A random sample $\{(X_i, Y_i)\}_{i=1}^n$ is drawn from this population, and data points are ranked in the ascending order of X. The n units are divided into g groups (called fractile groups) each of equal size $n' = (n/g)$. The mean of the $y-$variable in each fractile group is calculated and labeled as $y'_1, y'_2, \ldots, y'_g$. g equidistant points $1, 2, \ldots, g$ on the $x-$axis are marked to represent the g fractile groups and the corresponding values of $y'_1, y'_2, \ldots, y'_g$ are plotted. Each pair of adjoining points y'_i with y'_{i+1} for $i = 1, 2, \ldots, g-1$ are joined by straight lines to get a polygonal curve called the fractile graph. Since the estimate consists of continuously joined straight line segments, the estimate is usually not smooth even though the population fractile regression function may be smooth in nature.

Though our estimation procedure is partly on the lines suggested by Mahalanobis, it is intrinsically different from his approach both methodologically and conceptually. In fact, our method blends Mahalanobis' ideas with more recent nonparametric kernel regression procedures; see e.g., Muller (1988), Härdle (1990), Wand and Jones (1995), Fan and Gijbels (1996) and Simonoff (1996). We have used smooth kernel regression estimators to form the estimated fractile graphs whereas Mahalanobis took the y-averages from each fractile group and constructed the estimated fractile graphs.

Suppose that we have data $\{(X_i, Y_i)\}_{i=1}^n$ from a bivariate population. Throughout this paper, we consider smooth estimates of the fractile graph $\widehat{m}_n(\cdot)$ of the form

$$\widehat{m}_n(t) = \sum_{i=1}^{n} Y_i W_{n,i}(t, h_n), \tag{1}$$

where $W_{n,i}(t, h_n) = W_{n,i}(t, h_n, F_n(X_1), F_n(X_2), \ldots, F_n(X_n)), 1 \leq i \leq n$, is a weight function associated with the kernel smoother. Here F_n is the empirical distribution function of $\{X_i\}_{i=1}^n$ (to avoid extreme values of $F_n(x)$ we define $F_n(x) = \frac{1}{n+1} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x\}}$ for $x \in \mathbb{R}$), and h_n is the bandwidth based on a sample of size n. Note that as we are regressing Y on the quantiles of X, in a sense, we can pretend that our observations are $\{(F(X_i), Y_i)\}_{i=1}^n$, where F is the distribution function of X_1. But as the distribution function F is not known, we work with the empirical distribution function F_n, and it is used in the weight functions for the fractile graph estimators. Some common choices for the weight functions available in kernel regression literature are described below.

1. Nadaraya-Watson type weight function:

   $$W_{n,i}(t, h_n) = \frac{K\left(\frac{t - F_n(X_i)}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{t - F_n(X_i)}{h_n}\right)},$$

   $i = 1, 2, \ldots, n$; where K is the kernel (can be the standard normal density function) and F_n is the empirical distribution function (see Nadaraya, 1964 and Watson, 1964).

2. Priestley-Chao type weight function: $W_{n,i}(t, h_n) = \frac{1}{nh_n} K\left(\frac{t - F_n(X_i)}{h_n}\right)$, $i = 1, 2, \ldots, n$ (see Priestly and Chao, 1972).

3. Gasser-Müller type weight function: $W_{n,(i)}(t, h_n) = \int_{t_{i-1}}^{t_i} \frac{1}{h_n} K\left(\frac{t-s}{h_n}\right) ds$, i=1, 2, …, n; where $0 = t_0 < \frac{1}{n+1} < t_1 < \frac{2}{n+1} < t_2 < \ldots < \frac{n}{n+1} < t_n = 1$ (see Muller, 1998).

4. Local linear type weight function:

$$W_{n,i}(t, h_n) = \frac{1}{n} \frac{\{\widehat{s}_2(t, h_n) - \widehat{s}_1(t, h_n)(F_n(X_i) - t)\} K\left(\frac{t - F_n(X_i)}{h_n}\right)}{\widehat{s}_2(t, h_n)\widehat{s}_0(t, h_n) - \widehat{s}_1(t, h_n)^2},$$

$i = 1, 2, \ldots, n$; where $\widehat{s}_r(t, h_n) = \frac{1}{nh_n} \sum_{i=1}^{n} (F_n(X_i) - t)^r K\left(\frac{t - F_n(X_i)}{h_n}\right)$ for $r = 0, 1, 2$ (see Wand and Jones, 1995, Fan and Gijbels, 1996) .

The choice of the bandwidth h_n is of crucial importance for the above estimators. For choosing the optimal bandwidth for the kernel regression function, we can use any standard data-driven bandwidth selection procedure applicable to kernel regression like the least squares cross validation (see Rice, 1984, Härdle, 1990, Wand and Jones, 1995, Fan and Gijbels, 1996) or direct plug-in method (see Wand and Jones, 1995).

As an example of fractile graphs, we demonstrate our smooth estimates of fractile graphs along with Mahalanobis' fractile graphs in Figure 1. Here we have generated two samples, each of size 100, from the population $y = 1.0 + x + \epsilon$ where $\epsilon \sim N(0, 0.09)$ and $x \sim Exp(1)$ (the first sample is represented by dots while the other one is represented by asterisks). The covariate for the second sample is then squared. Note that the scatter plots of the two samples look very different due to the square transformation on the covariate for the second population. But after transforming the covariate (i.e., changing X_i to $F_n(X_i)$) the scatter plots look quite similar. Thus Mahalanobis' fractile graphs and our smooth estimated fractile graphs are quite similar in nature. For our smooth estimated fractile graphs we have used the Nadaraya-Watson type weight function with standard normal kernel and least squares cross validated optimal bandwidths. The chosen bandwidths for the two populations are 0.027 and 0.022 respectively. As the bandwidths are quite small, the estimated smooth fractile graphs are wiggled in nature.

Although we have used smooth kernel based estimators to estimate the fractile graphs, the setup discussed in the last section is very general. A large class of nonparametric regression estimators can be expressed in the form (1). In the next section we present asymptotic properties of the estimated fractile graphs that are valid in the general setup under appropriate conditions on the weight functions. We have used kernel smoothing in the data analytic section for its simplicity and computational ease. Asymptotics also become tractable as the weight functions can be easily computed. Later on, while comparing two fractile graphs, we adopt the multi-scale methodology where results are available for kernel type weight functions (see Chaudhuri and Marron, 2000).

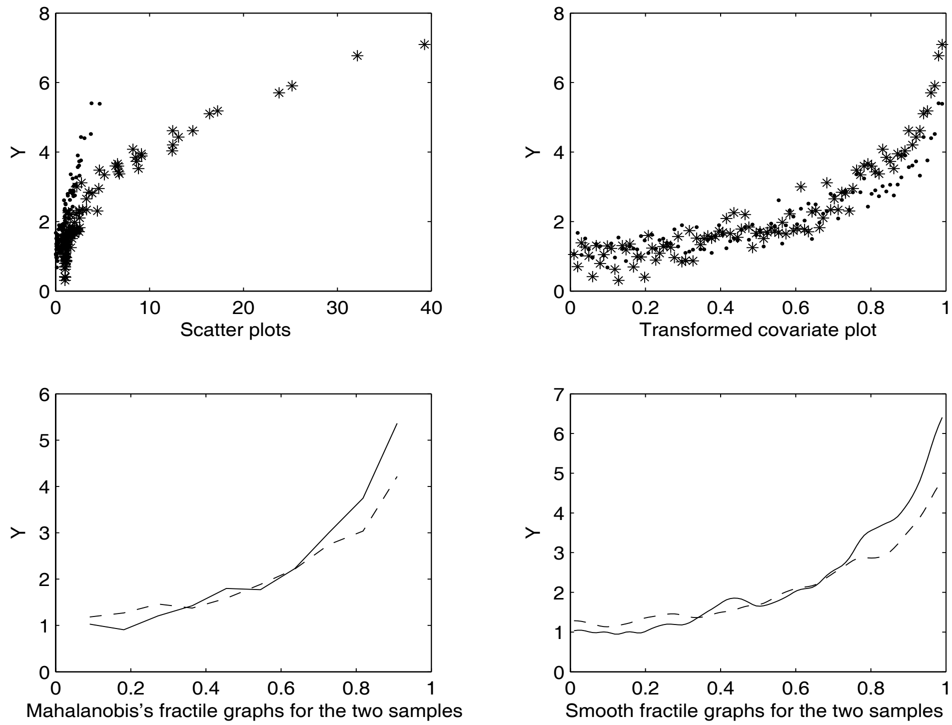

Figure 1. Scatter plots for the two samples (top left), Transformed covariate scatter plots (top right), Mahalanobis' fractile graphs with 10 fractile groups (bottom left) and smooth estimates of fractile graphs with bandwidths chosen by least squares cross-validation (bottom right).

Another way to estimate the smooth fractile graphs could be by computing the usual nonparametric estimator $\widehat{g}_n(x)$ of $g(x) = E(Y|X = x)$ and evaluating this at $x = G_n^{-1}(t)$ where $G_n^{-1}(t)$ is the $t-$th sample quantile of X. We would then require a strictly increasing estimator of G_n (the usual empirical distribution function of X will not suffice) and that would have to be inverted to get the corresponding quantile. All this would make the estimation of fractile graphs much more computationally intensive. We do not discuss this approach in this paper.

3 Asymptotic Properties of Smooth Fractile Regression Estimate

Bhattacharya and Muller (1993), Parthasarathy and Bhattacharya (1961), and Sethuraman (1961) studied in detail the asymptotic properties of Mahalanobis' original estimates of fractile graphs as the sample size grows but the number of fractile groups are kept fixed. In the first part of this section, we prove the consistency of our smooth estimates of fractile graphs under mild regularity conditions (stated below). In the later part of the section, we prove the asymptotic normality of the smooth estimates of fractile graphs. Throughout this section we will assume that the covariate $X \sim F$ where F is a continuous strictly increasing distribution function. We begin by stating some conditions that will be needed to establish consistency of kernel estimates of fractile regression.

(C1) The function $m(\cdot)$ is continuous $\forall\, t \in (0,1)$, and $m(t) < M \;\forall\, t \in [0,1]$ for some $M > 0$.

(D1) The conditional variance function $v(t) = \text{Var}\,(Y|F(X) = t)$ is bounded above by $K_0 > 0$ (i.e., $v(t) \leq K_0 \;\forall\, t \in [0,1]$).

(D1′) There exists $B > 0$ such that $|Y| < B$ almost surely.

(W1) $\sum_{i=1}^{n} W_{n,i}^2(t, h_n) \longrightarrow 0$.

(W2) $\sum_{i=1}^{n} W_{n,i}(t, h_n) \longrightarrow 1$.

(W3) Given any $\rho > 0$, $\exists\, A > 0$ and $N \in \mathbb{N}$ such that

$$\sum_{i=1}^{n} |W_{n,i}(t, h_n)| \mathbf{1}_{\left\{\left|\frac{t - F_n(X_i)}{h_n}\right| > A\right\}} < \rho \;\forall\, n \geq N.$$

(W4) There exists $C > 0$, such that $\sum_{i=1}^{n} |W_{n,i}(t, h_n)| \leq C \;\forall\, n \geq 1$.

An interesting point to be observed here is that the sums appearing in conditions (W1)–(W4) are not dependent on the observations $X_1, X_2, \ldots, X_n$, as the weights depend only on $F_n(X_i)$'s and not on X_i's.

THEOREM 3.1 *Fix $0 < t < 1$. Let us assume (C1) and (W1)–(W4). Then, under assumption (D1), the conditional MSE of $\widehat{m}_n(t)$ given the X_i's tend to 0 almost surely as $n \to \infty$. As a consequence,*

$$\widehat{m}_n(t) \xrightarrow{P} m(t) \text{ as } n \to \infty. \tag{2}$$

Further, if (D1) is strengthened to (D1′), we have

$$\int \{\widehat{m}_n(t) - m(t)\}^2 dt \xrightarrow{P} 0 \text{ as } n \to \infty. \tag{3}$$

Let us next consider following conditions:

(C1′) $m(\cdot)$ is twice continuously differentiable on the interval $[0, 1]$.

(K1) The kernel function $K(\cdot)$ is symmetric about 0 with a bounded derivative, nonzero in a neighbourhood of 0, and it has compact support (i.e., $K(x) = 0\ \forall\ |x| \geq M_0$ for some $M_0 > 0$). The bandwidth sequence associated with the kernel smoother satisfies $\lim_{n\to\infty} h_n = 0$ and $\lim_{n\to\infty} nh_n = \infty$.

(K2) The kernel function $K(\cdot)$ is symmetric about 0 with a bounded derivative, nonzero in a neighbourhood of 0, and it is supported on the entire real line. The bandwidth sequence associated with the kernel smoother satisfies $\lim_{n\to\infty} h_n = 0$ and $\lim_{n\to\infty} nh_n^2 = \infty$.

Corollary 3.2 *Assume that either (K1) or (K2) holds. Then, under conditions (C1) and (D1), (2) holds for estimates based on Nadaraya-Watson, Priestley-Chao, and Gasser-Muller type weight functions. If (D1) is strengthened to (D1′), (3) holds for estimates based on the same weight functions. On the other hand, under conditions (C1′), (D1) and either of (K1) or (K2), we have (2) for estimates based on local linear type weight function. Further, if (D1) is replaced by (D1′), (3) holds for local linear estimates of fractile graphs.*

To prove the asymptotic normality of the estimated fractile graph we need slightly stronger assumptions. We state the following conditions which will be used to prove the next theorem.

($\tilde{C}1$) The function $m(\cdot)$ is differentiable $\forall\ t \in (0, 1)$, and $m'(t) < K\ \forall\ t \in [0, 1]$ for some $K > 0$.

($\tilde{W}1$) $\frac{\sum_{i=1}^n W_{n,i}(t,h_n)-1}{\sqrt{\sum_{i=1}^n W_{n,i}^2(t,h_n)}} \longrightarrow 0$ as $n \to \infty$.

($\tilde{W}2$) There exists $A > 0$ such that $\sum_{i=1}^n |W_{n,i}(t, h_n)| \mathbf{1}_{\{\left|\frac{t-F_n(X_i)}{h_n}\right|>A\}} = 0$.

($\tilde{W}3$) $\frac{\sum_{i=1}^n W_{n,i}^2(t,h_n)}{\max_{1\leq i\leq n} W_{n,i}^2(t,h_n)} \longrightarrow \infty$ as $n \to \infty$.

($\tilde{W}4$) $nh_n^3 \longrightarrow 0$ as $n \to \infty$.

Theorem 3.3 *Fix $0 < t < 1$. Let us assume $(\tilde{C}1)$ and $(\tilde{W}1) - (\tilde{W}4)$. Suppose that $Y_i = g(X_i) + \epsilon_i, i = 1, 2, \ldots, n$; where $\epsilon_i \perp X_i, i = 1, 2, \ldots, n$, and $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2$. Then,*

$$\frac{\widehat{m}_n(t) - m(t)}{\sigma\sqrt{\sum_{i=1}^{n} W_{n,i}^2(t, h_n)}} \xrightarrow{d} N(0,1) \text{ as } n \to \infty. \tag{4}$$

Corollary 3.4 *Suppose that the kernel function $K(\cdot)$ is symmetric about 0 with a bounded derivative, nonzero in a neighbourhood of 0, and has compact support (i.e., $K(x) = 0 \ \forall \ |x| \geq M_0$ for some $M_0 > 0$). Also suppose that the bandwidth sequence associated with the kernel smoother satisfies $\lim_{n\to\infty} h_n = 0$ and $\lim_{n\to\infty} nh_n^3 = \infty$. Then (4) holds for estimates based on Nadaraya-Watson and Gasser-Muller type weight functions.*

4 Comparison of Estimated Smooth Fractile Graphs

Suppose that we have data from two populations: $\{(X_{1i}, Y_{1i})\}_{i=1}^{n_1}$ and $\{(X_{2i}, Y_{2i})\}_{i=1}^{n_2}$, and we want to test the hypothesis

$$H_0 : m_1 = m_2 \qquad vs. \qquad H_A : m_1 \neq m_2, \tag{5}$$

where $m_1(t) = E(Y_{1i}|F_1(X_{1i}) = t)$ and $m_2(t) = E(Y_{2i}|F_2(X_{2i}) = t)$, and F_1 and F_2 are continuous strictly increasing distribution functions of the random variables X_{1i} and X_{2i}.

Much effort has been devoted to the problem of comparison of nonparametric regression curves in the recent literature; see e.g., Härdle and Marron (1990), King, Hart and Wehrly (1991), Hall and Hart (1990), Delgado (1993), Young and Bowman (1995), Hall, Huber and Speckman (1997), Munk and Dette (1998), Dette and Neumeyer (2003). These authors considered the testing problem

$$H_0 : g_1 = g_2 \qquad vs. \qquad H_A : g_1 \neq g_2, \tag{6}$$

where g_1 and g_2 are the usual regression curves corresponding to two different populations. Most authors concentrated on equal design points to develop tests for (6). Hall and Hart (1990) discussed a bootstrapping procedure for testing (6) under the assumption of common design points and same number of data points for both the samples. They extended their tests to samples with unequal design points but with the same sample size. Härdle and Marron (1990) suggested a semiparametric method of testing (6) with equal design points. They parameterize the difference between the two curves. King, Hart and Wehrly (1991) proposed two tests for common design points: one for normal errors and the other for non-normal errors that are shown to have good power properties. Kulasekera (1995) proposed a test for the hypothesis (6) using quasi-residuals which is applicable under the assumption of different design points for both the samples. Kulasekera and Wang (1997) considered the selection of smoothing parameters to obtain optimal power in tests of regression curves. Munk and Dette (1998), Neumeyer and Dette (2003) considered the problem of the comparison of nonparametric regression curves under a very general set-up. Delgado (1993), Kulasekera (1995) and Kulasekera and Wang (1997) considered marked empirical processes to develop tests for the hypothesis (6). However, none of the above mentioned authors has addressed the problem of possible effects of some transformation on the covariate for the two populations. Some of the usual methods for comparison of the regression curves do not generalize in a straight forward manner in our setup as in fractile regression the covariate X_i is replaced by $F_n(X_i)$, and the $F_n(X_i)$'s are not independent even if the X_i's are so. We next describe Mahalanobis' original approach towards this problem.

4.1. Mahalanobis' idea for comparing two fractile graphs. The first sample of size n_1 is obtained from the first bivariate population by drawing two independent ("interpenetrating") random half-samples each of size $n_1/2$. The first half-sample is then considered, and the fractile graph $G(1)$ is constructed from it [see Section (2) for the construction of Mahalanobis' fractile graphs]. The second half-sample is used to get the second fractile graph $G(2)$. Clearly, the two half-sample fractile graphs $G(1)$ and $G(2)$ have identical statistical distributions.

Mahalanobis' idea was to mix the two half-samples to form the combined sample of size n_1 from the first population. The combined sample is again ranked according to the X-values and divided into g fractile groups each containing n_1' ($n_1' = n_1/g$) units. The y-averages of the corresponding fractile groups are plotted to get the combined fractile graph $G(1,2)$. The "error area" $a(1,2)$ associated with the combined sample is defined as the area bounded between the two half-sample fractile graphs $G(1)$ and $G(2)$ (i.e., $a(1,2) = \int |G(1) - G(2)|$).

The second bivariate population is considered next from which a pair of independent ("interpenetrating") half-samples are drawn. The second set of fractile graphs $G'(1), G'(2)$ and $G'(1,2)$ are computed from the half-samples obtained from the second population. The area bounded between $G'(1)$ and $G'(2)$ is called the second "error area" associated with the second population and is denoted by $a'(1,2)$ (i.e., $a'(1,2) = \int |G'(1) - G'(2)|$). The area between the two combined fractile graphs $G(1,2)$ and $G'(1,2)$ is called the "separation area" between $G(1,2)$ and $G'(1,2)$ and is denoted by $S(1,2)$ (i.e., $S(1,2) = \int |G(1,2) - G'(1,2)|$).

The statistical error E to be associated with the "separation area" $S(1,2)$ is defined by the formula $E = \sqrt{a^2(1,2) + a'^2(1,2)}$. The significance of the observed value of $S(1,2)$ is tested by considering the test-statistic $\frac{S^2(1,2)}{E^2}$, which according to Mahalanobis would be distributed approximately like a chi-square random variable. For some of the statistical properties of the "error area" in FGA see Takeuchi (1961), Mitrofanova (1961), Mahalanobis (1988).

4.2. Statistical comparison of smooth estimates of fractile graphs. As mentioned above, we use smooth estimates of fractile graphs $\widehat{m}_{1,n_1}$ and $\widehat{m}_{2,n_2}$. To test hypothesis (5), we will use the following test statistic

$$T_{n_1,n_2} = \int_0^1 |\widehat{m}_{1,n_1}(t) - \widehat{m}_{2,n_2}(t)| dt \tag{7}$$

which we call as the "separation area" between the smooth fractile graphs. We have also considered another test statistic defined as

$$S_{n_1,n_2} = \int_0^1 \{\widehat{m}_{1,n_1}(t) - \widehat{m}_{2,n_2}(t)\}^2 dt \tag{8}$$

and call it the "squared difference between the fractile graphs". Under the null hypothesis, we expect the test statistics to be small, whereas large values of the test statistics would support the alternative hypothesis.

We have adopted the multi-scale approach to curve estimation, where one considers a range of possible bandwidths and statistical tests are carried out for all of them, as e.g., in SiZer (see Chaudhuri and Marron, 1999, 2000). In this approach, we shift our attention from the "true underlying curve" to the "true curves viewed at different levels of smoothing" with the idea that it will enable us to extract relevant information available in the noisy data at different levels of smoothing. This approach is more informative than an approach based on a single "optimal" bandwidth. This approach is an alternative to selecting an optimal data-driven bandwidth. The poor performance of some of the data-driven bandwidth selection procedures have led us adopt this multi-scale approach to curve estimation.

4.3. Resampling based test using T_{n_1,n_2} and S_{n_1,n_2}. A major technical barrier using the test statistics T_{n_1,n_2} and S_{n_1,n_2} [defined in (7) and (8)] is that their sampling distributions are analytically intractable. In this section we provide detailed procedures to test the hypothesis (5) based on these test statistics using resampling techniques.

4.3.1. Bootstrap method. Bootstrap (see e.g., Hall, 1992 and Efron and Tibshirani, 1993) is a very well known and widely used resampling procedure, which is used to estimate the distribution of a test statistic for which the sampling distribution is otherwise intractable. We describe below the steps involved in computing the boostrap estimates of the P-values when T_{n_1,n_2} and S_{n_1,n_2} are used as test statistics.

- We transform the covariate X into its quantiles, i.e., we transform the data set into $\{(\frac{i}{n_1+1}, Y_{1[i:n_1]})\}_{i=1}^{n_1}$ and $\{(\frac{i}{n_2+1}, Y_{2[i:n_2]})\}_{i=1}^{n_2}$, where $Y_{1[i:n_1]}$ denotes the concomitant of the i-th order statistic of X for the first sample (see e.g., David and Nagaraja, 1998, 2003 and Yang, 1977). These transformed covariate for the two samples help us compute the weight functions used in the smooth estimation of fractile graphs.

- After transforming the covariate, we obtain the smooth fractile graphs for the two populations as explained in Section 2. We compute T_{n_1,n_2} and S_{n_1,n_2} from the data. We use the multi-scale approach discussed above, and fix bandwidths for the kernel smoother at different values for the two samples to carry out our analysis for each chosen pair.

- To test the significance of the observed values of T_{n_1,n_2} and S_{n_1,n_2}, we bootstrap from the joint distribution of (X, Y), i.e., from the joint density of the original data sets. We resample data sets from a kernel estimate of the joint density of (X, Y).

  One could have bootstrapped from the empirical joint distribution but that would result in repetitions of data-points and would distort the fractile graphs. This is because the empirical distribution function is not continuous. We separately compute the joint densities, using usual kernel density estimators, for the two populations as described below

  $$f_{X,Y}(x, y) = \frac{1}{n}\frac{1}{h_x h_y}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_x}\right) K\left(\frac{y - Y_i}{h_y}\right) \ \forall\ (x, y) \in \mathbb{R}^2,$$

  where K is a symmetric and bounded kernel.

- For the kernel density estimators, we have to decide on the choice of the bandwidth; when standard cross-validation methods were tried out, it did not work well in our numerical investigations. Further, it is computationally very expensive. After some empirical investigations, we decided to use kernel density estimates with fixed "very small" bandwidths for both the variables, namely $h_x = c_1/n$ and $h_y = c_2/n$, where c_1 and c_2 are the standard deviations of X and Y respectively. This "undersmoothed" density estimate leads to a distribution estimate which is very close to the empirical distribution function. However, the distribution estimate is now continuous and strictly increasing – something that we need for proper estimates of the fractile graphs.

- We draw two independent bootstrap samples of size n_1 and n_2 from the estimated density of the first population, and let us denote it by $\{(\tilde{X}_{1i}, \tilde{Y}_{1i})\}_{i=1}^{n_1}$ and $\{(\tilde{X}_{2i}, \tilde{Y}_{2i})\}_{i=1}^{n_2}$. Let $\tilde{T}^{(1)}$ and $\tilde{S}^{(1)}$ be the "error area" and "squared difference between the two bootstrapped fractile graphs", respectively, computed from those bootstapped samples. Let $\tilde{T}^{(2)}$ and $\tilde{S}^{(2)}$ be defined similarly for the bootstrapped fractile graphs obtained from two independent bootstrapped samples drawn from estimated density of the second population. These computations are repeated N times (in our numerical studies we have used $N = 1000$) to yield $\{\tilde{T}_i^{(1)}\}_{i=1}^{N}$, $\{\tilde{T}_i^{(2)}\}_{i=1}^{N}$, $\{\tilde{S}_i^{(1)}\}_{i=1}^{N}$ and $\{\tilde{S}_i^{(2)}\}_{i=1}^{N}$.

- Note that T_{n_1,n_2} and S_{n_1,n_2} measure the difference between the two estimated fractile graphs. To test the equality of the fractile graphs we compare this observed difference between the fractile graphs (T_{n_1,n_2} and S_{n_1,n_2}) with the distribution of the test statistics when the two population fractile graphs are actually equal. Thus, under the null hypothesis, the observed values of T_{n_1,n_2} and S_{n_1,n_2} would be comparable to the distributions of $\tilde{T}^{(1)}$ and $\tilde{S}^{(1)}$ and also that of $\tilde{T}^{(2)}$ and $\tilde{S}^{(2)}$ respectively. Comparing T_{n_1,n_2} and S_{n_1,n_2} with the distributions of $\tilde{T}^{(1)}$ and $\tilde{S}^{(1)}$ would be like taking the first population as the null hypothesis population. The bootstrap estimates of the P-values for this comparison are the proportion of times $\tilde{T}_i^{(1)}$ exceeds the observed value of T_{n_1,n_2} and the proportion of times $\tilde{S}_i^{(1)}$ exceeds the observed value of S_{n_1,n_2}. Similarly, if the second population is taken as the null hypothesis population (i.e., we compare the observed values of the test statistics to the distribution of the test statistics obtained from the second population), the P-values are computed in the same manner using $\tilde{T}_i^{(2)}$ and $\tilde{S}_i^{(2)}$. It is like the centering that is needed in a bootstrap test to force the null hypothesis to be correct. Note that here for each test statistic we get two bootstrap estimates of the P-value.

4.3.2. Swap method. In this section, we propose another method that is computationally faster than the bootstrap and yields only one P-value for a given test-statistic. The steps involved in this procedure are described below.

- Once again, we transform the covariate (i.e., X) into its corresponding quantiles.

- After transforming the covariate, we construct the smooth estimates of fractile graphs and compute the statistics T_{n_1,n_2} and S_{n_1,n_2}. Here also we have used the multi-scale analysis. We fix the kernel bandwidths at different values for both the populations and carry out our analysis separately for each pair of bandwidths chosen for the two populations as before.

- Assume next that $n_1 = n_2$. To test the significance of observed values of T_{n_1,n_2} and S_{n_1,n_2}, we resample a pair of data sets from the original transformed data sets in the following way. We swap (or interchange) the i-th ranked data point (i.e., the data point $(\frac{i}{n_1+1}, Y_{1[i:n_1]})$) of the first population with the i-th ranked data point of the second population (i.e., the data point $(\frac{i}{n_2+1}, Y_{2[i:n_2]})$) with probability 0.5 and keep it unchanged with probability 0.5. This mixes the two data sets accordingly. Using the resampled data sets we re-calculate the fractile graphs and the corresponding test statistics.

- When $n_1 \neq n_2$, the i-th ranked data points for the two samples might not correspond to the same sample quantile. In this case, one can interpolate between covariate quantiles along with the corresponding y-values before carrying out the swap operation.

- The swap operation leads to resampled data sets,i.e., $\{(X_{1i}^{(*)}, Y_{1i}^{(*)})\}_{i=1}^{n_1}$ and $\{(X_{2i}^{(*)}, Y_{2i}^{(*)})\}_{i=1}^{n_2}$. These resampled data sets are used to compute $T^* =$ the "resampled separation area" and $S^* =$ the "resampled squared difference between resampled fractile graphs".

- We repeat the entire procedure N times as in the case of bootstrap (here also we use $N = 1000$ for our numerical studies) to get $\{T_i^*\}_{i=1}^N$ and $\{S_i^*\}_{i=1}^N$. Finally the P-values corresponding to the two test statistics are obtained as the proportion of times T_i^* exceeds the observed value of T_{n_1,n_2} and the proportion of times S_i^* exceeds the observed value of S_{n_1,n_2}.

5 Data Analysis

In the first part of this section we illustrate the performance of the testing procedures on synthetic data. We analyze some data collected from the Reserve Bank of India in the second part of the section.

As mentioned earlier, we use the multi-scale approach to test significant difference in a pair of smooth estimates of fractile graphs. We plot the P-values obtained from the multiple tests against the choice of the two bandwidths of the kernel smoothed fractile graphs for the two populations. This gives us a two-dimensional plot over a rectangle (please see Figures 5, 7 and 9), where high P-values correspond to white regions and low P-values correspond to dark regions. We draw inference about the two populations based on the grey-scale plots obtained by the above procedure. We have used the Nadaraya-Watson type weight function with the standard normal kernel in all the examples considered below. We have taken a broad range of bandwidths for the grey-scale plots varying from 0.01 to 0.40. The minimum bandwidth is chosen so that at least two data-points have significant contributions in the estimated fractile graph at each point. The maximum value of the bandwidth is chosen in such a manner that all the data-points have some contribution in the estimate at each point. These values can be computed using the fact that the standard normal kernel is practically zero outside $(-3, 3)$.

The two methodologies considered by us for comparison of the fractile graphs are :

1. Resample from the joint density of (X,Y) (bootstrap method) and

2. Swap the response for the same fractile value (swap method).

5.1. Evaluation of the resampling based tests using synthetic data. In this section we present a small-scale simulation study of the multi-scale analysis to investigate the P-values in cases when the null hypothesis is true and also when the null hypothesis is false. The examples illustrate the small sample behaviour of the resampling methods. Throughout the entire section we work with two samples from two populations with sample sizes $n_1 = n_2 = 100$. We have used $l(= 500 \text{ or } 1000)$ iterations to compute the P-values. We plot the grey-scale images of the P-values computed using the test statistics (i) T (absolute difference) and (ii) S (squared difference between the fractile graphs) separately.

Example 1. We generate two samples, each of size 100, from the population $y = 1.0 + x + \epsilon$ where $\epsilon \sim N(0, 0.09)$ and $x \sim Exp(1)$. The covariate for the second sample is then squared. This gives us the two samples for comparison. Thus, the population fractile graphs are the same for both the populations. We replicate this generation method 10 times and plot the average P-value for each pair of bandwidth in the following grey-scale plots (see Figure 2). The P-values are quite high varying mostly between 0.35 and 0.70 for the bootstrap method and between 0.15 and 0.60 for the swap method.

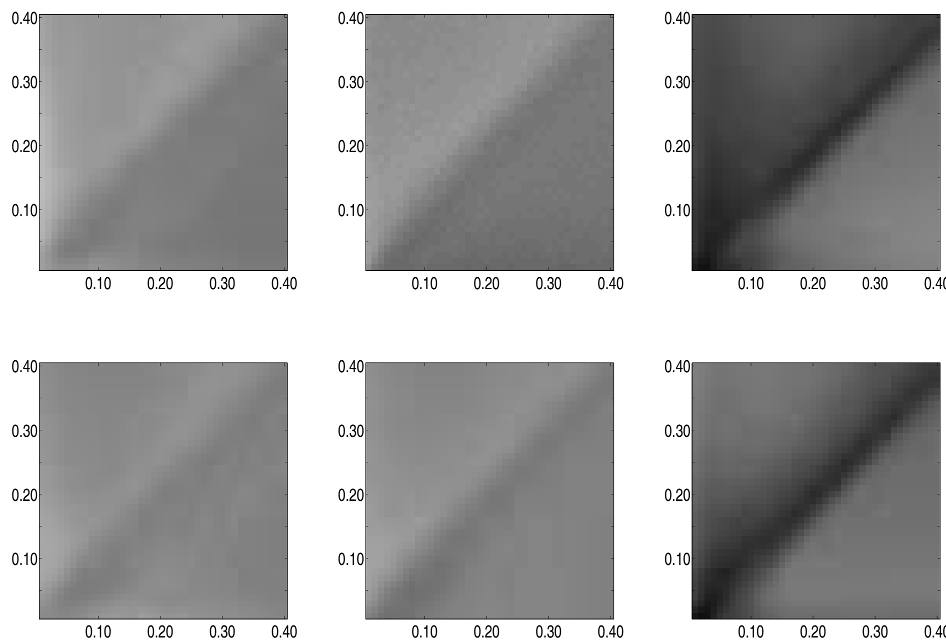

Figure 2. Grey-scale images of the P-values for Example 1, where in each frame the bandwidths corresponding to the first sample are plotted along the horizontal axis, and those corresponding to the second sample are plotted along the vertical axis. The top and the bottom rows correspond to the test statistics T_{n_1,n_2} and S_{n_1,n_2} respectively. The first two columns are images of bootstrapped P-values obtained by using the first and second population as the null hypothesis populations, respectively. The third column gives images of P-values obtained by the swap method.

EXAMPLE 2. In this example, we generate samples from different models. The first sample is generated from the population $y = 1.0 + x + \epsilon$ where $\epsilon \sim N(0, 0.09)$ and $x \sim Exp(1)$. The second sample is from the population $y = 1.3 + x + \epsilon$ where $\epsilon \sim N(0, 0.09)$ and $x \sim Exp(1)$. The covariate for the second sample is again squared. The fractile graphs for the two populations are different. As in the previous example, we replicate this generation method 10 times and plot the average P-value for each pair of bandwidth in Figure 3. The P-values are very close to zero, mostly ranging from 0.00 to 0.15 for the bootstrap method and from 0.00 to 0.10 for the swap method.

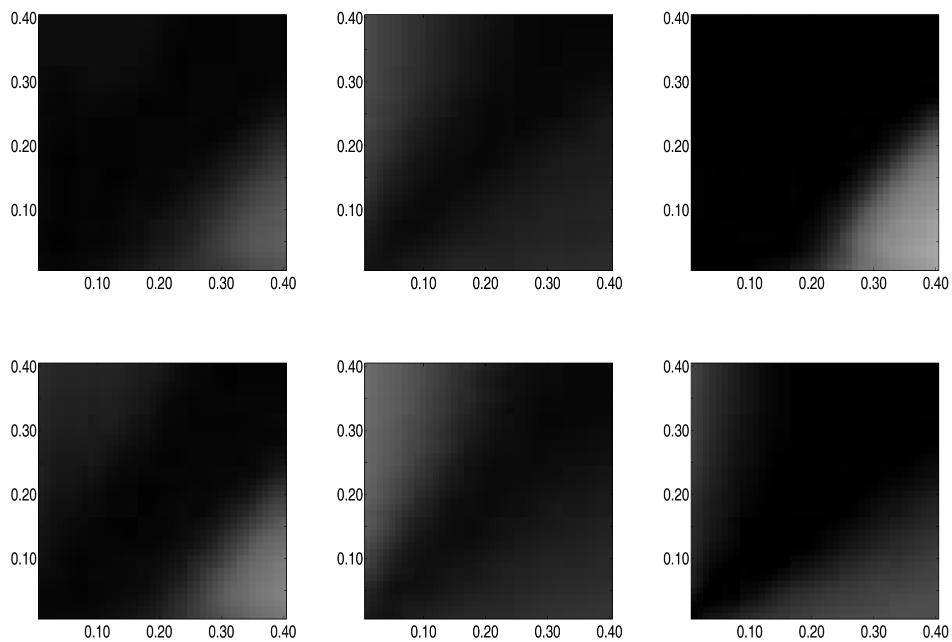

Figure 3. Grey-scale images of the P-values for Example 2, where in each frame the bandwidths corresponding to the first sample are plotted along the horizontal axis, and those corresponding to the second sample are plotted along the vertical axis. The top and the bottom rows correspond to the test statistics T_{n_1,n_2} and S_{n_1,n_2} respectively. The first two columns are images of bootstrapped P-values obtained by using the first and second population as the null hypothesis populations, respectively. The third column gives images of P-values obtained by the swap method.

5.2. Some real data illustrations.

EXAMPLE 3. *Credit-Deposit Ratio vs. Total Deposit.* We obtained the data from "Basic Statistical Report of the Scheduled Commercial Banks in India" published annually by the Reserve Bank of India (refer to the link "Basic Statistical Report of the Scheduled Commercial Banks in India" at the website http://www.rbi.org.in). We collected statewise data from the 32 states in India (excluding Jharkhand, Chattisgarh, and Uttaranchal) for all Scheduled Commercial Banks in India on Credit-Deposit Ratio and Total Deposit for the years 1996 and 2002. We regress Y = Credit-Deposit Ratio = Total Credit/Total Deposit on X = Total Deposit. Among the states, there are wide fluctuations in the value of Y in each year. In most years, the ratio Y varies from 0.2 to 1.4. The fractile graphs for the two samples corresponding to the years 1996 and 2002 are shown in Figure 4. The grey-scale plots for the comparison of the fractile graphs for the two samples are presented in Figure 5. The P-values are quite high, and they mostly vary between 0.65 to 0.95 indicating no statistically significant change from 1996 to 2002.

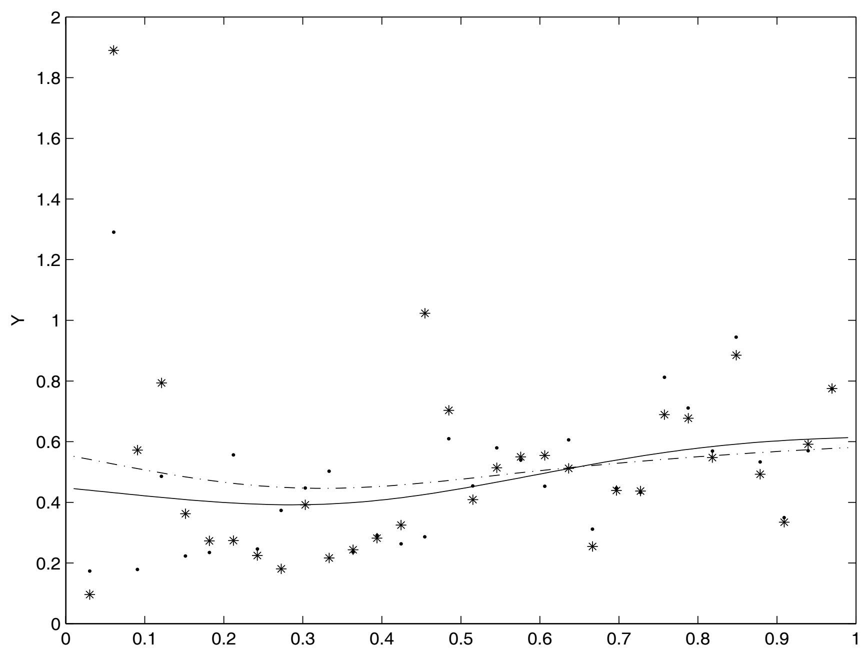

Figure 4. Smooth fractile graphs along with scatter plots for Credit-Deposit Ratio and Total Deposit for the years 1996 and 2002 using optimum least squares cross validation bandwidth (0.16 and 0.20 respectively).

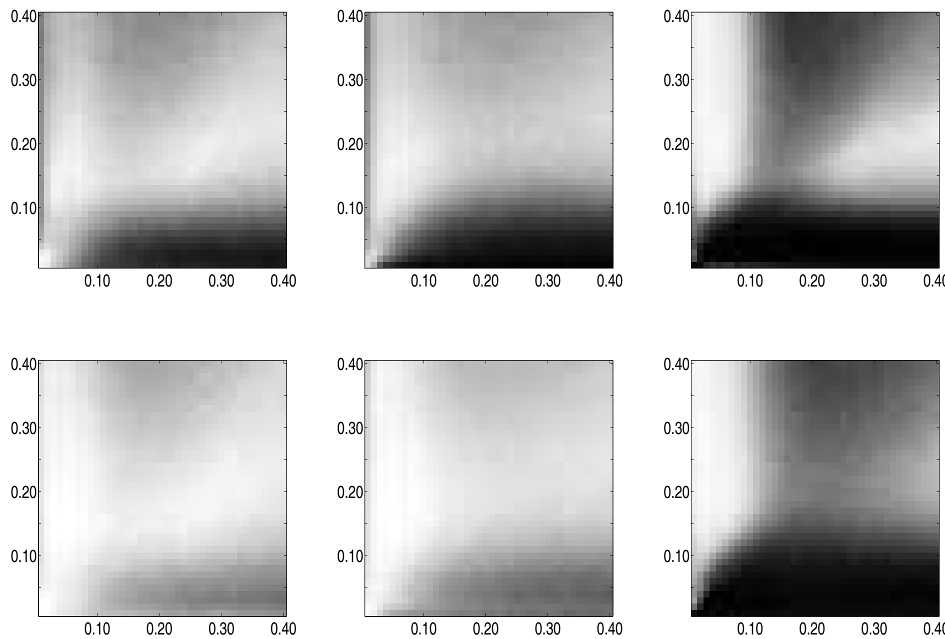

Figure 5. Grey-scale images of the P-values for Example 3 for comparing years 1996 with 2002, where in each frame the bandwidths corresponding to the estimate for the year 1996 are plotted along the horizontal axis, and those corresponding to the estimate for the year 2002 are plotted along the vertical axis. The top and the bottom rows correspond to the test statistics T_{n_1,n_2} and S_{n_1,n_2} respectively. The first two columns are images of bootstrapped P-values obtained by using the years 1996 and 2002 as the null hypothesis populations, respectively. The third column gives images of P-values obtained by the swap method.

Example 4. *Profit-to-Sales Ratio vs. Sales.* The Reserve Bank of India also maintains data on the annual abridged financial results of non-government, non-financial public limited companies over different years. The data consists of Sales and Profit-to-Sales, among many other variables related to the companies. The Reserve Bank of India is interested in comparing the performance and profitability of the companies of different sizes, over the years. An appropriate size measure can be Sales (in 10000 rupees). In this example, we regress Y = ratio of Profit-to-Sales = Profit/Sales against X = Sales. We first studied data for the years 1997 and 2003. The sample sizes for the years 1997 and 2003 are 945 and 1267 respectively. Over the years, the Profit-to-Sales ratio has decreased. The fractile graphs for the two samples corresponding to the years 1997 and 2003 are shown in Figure 6. The least squares cross validated optimal bandwidth for the year 1997 is quite small which would make the smooth fractile graph very wiggled. Thus we use similar bandwidths (the bandwidth chosen was the least squares optimal bandwidth for the fractile graph for year 2003) for constructing both the fractile graphs. Grey-scale plots similar to those presented for Example 3 are also given. The P-values for the pairwise comparison of the smooth estimated fractile graphs for the years 1997 and 2003 are plotted in Figure 7. The P-values are almost always zero or very close to zero, giving rise to completely black images. This fact indicates significant changes from 1997 to 2003.

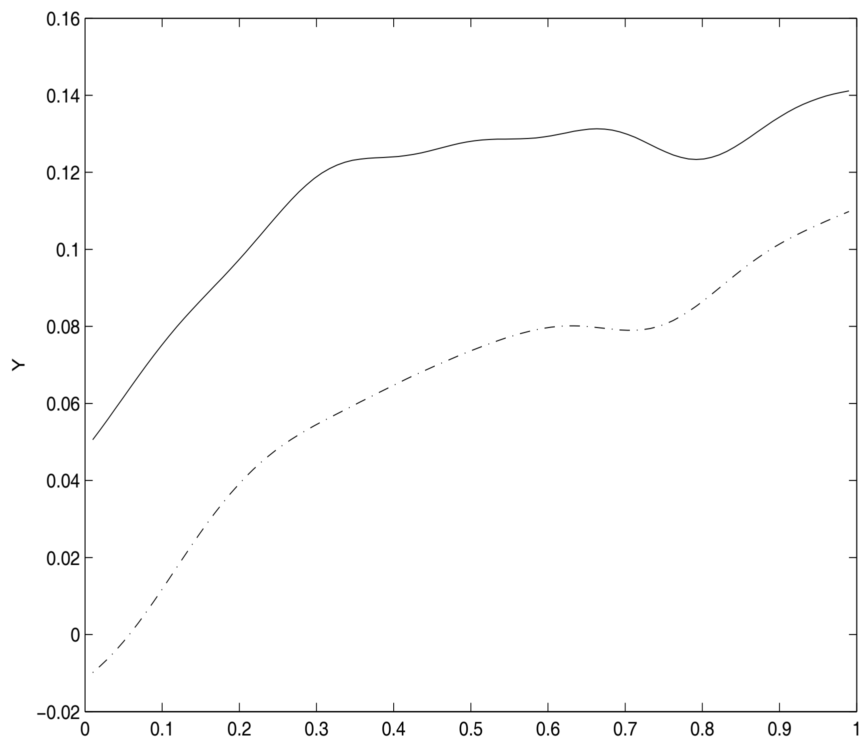

Figure 6. Smooth fractile graphs for Profit-to-Sales Ratio and Sales for the years 1997 and 2003 using same bandwidth (0.069 in both cases).

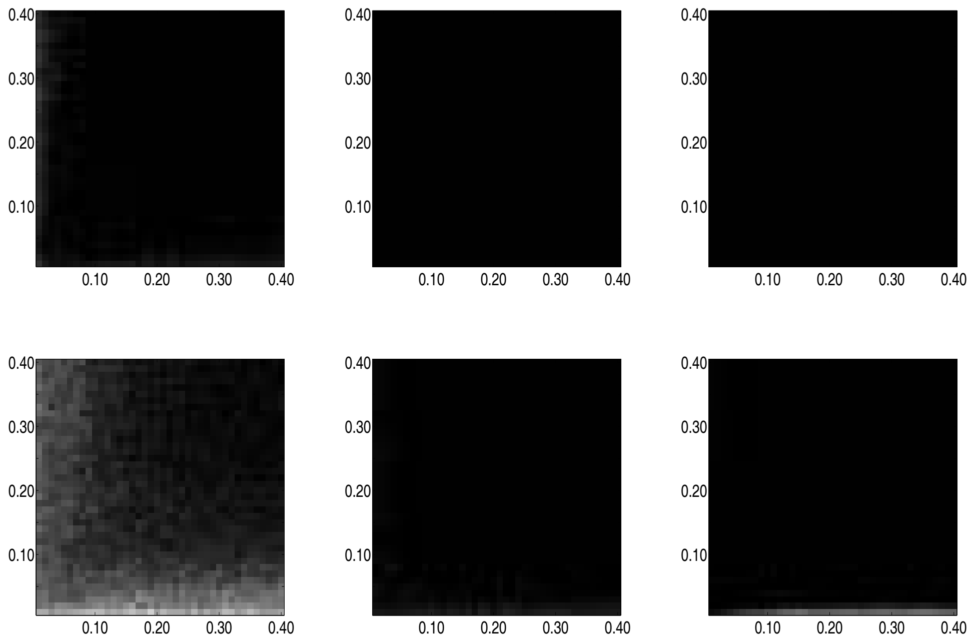

Figure 7. Grey-scale images of the P-values for Example 4 for comparing years 1997 with 2003, where in each frame the bandwidths corresponding to the estimate for the year 1997 are plotted along the horizontal axis, and those corresponding to the estimate for the year 2003 are plotted along the vertical axis. The top and the bottom rows correspond to the test statistics T_{n_1,n_2} and S_{n_1,n_2} respectively. The first two columns are images of bootstrapped P-values obtained by using the years 1997 and 2003 as the null hypothesis populations, respectively. The third column gives images of P-values obtained by the swap method.

As we have observed significantly small P-values for almost all choices of bandwidth pairs indicating significant change in the fractile graphs from 1997 to 2003, we decided to compare two consecutive recent years namely 2002 and 2003. The fractile graphs for the two samples are shown in Figure 8. Here again we observe that the optimal bandwidth for year 1997 is quite small and hence we use similar bandwidths for constructing both the fractile graphs. The P-value grey-scale plots are provided in Figure 9. In this case the P-values generally vary between 0.20 to 0.70. The minimum P-value observed is around 0.10 while the maximum is around 0.9. This is an indication that there has possibly been some changes from the year 2002 and 2003 but the evidence is statistically not very strong.

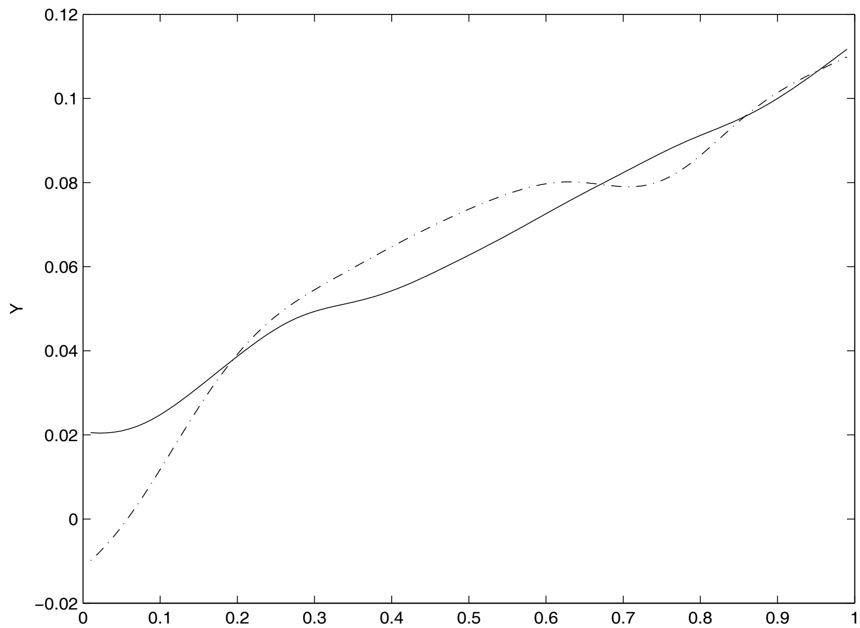

Figure 8. Smooth fractile graphs for Profit-to-Sales Ratio and Sales for the years 2002 and 2003 using same bandwidth (0.069 in both cases).

6 Proofs

PROOF OF THEOREM 3.1. We consider our data set as $\{(U_i, Y_i)\}_{i=1}^{n}$, where $U_i = F(X_i)$ is the transformed covariate. As the U_i's are not observed, we estimate them by $F_n(X_i)$'s and the weight functions depend on $F_n(X_i)$'s. Note that the conditional MSE given the U_i's can be decomposed as

$$E\{\widehat{m}_n(t) - m(t))\}^2 = E\{\widehat{m}_n(t) - E(\widehat{m}_n(t))\}^2 + \{E(\widehat{m}_n(t)) - m(t)\}^2. \quad (9)$$

Now, the conditional variance term can be simplified as

$$\begin{aligned} & E\left\{\widehat{m}_n(t) - E(\widehat{m}_n(t))\right\}^2 \\ = \;& E\left\{\sum_{i=1}^{n}(Y_i - E(Y_i|U_i))W_{n,i}(t,h_n)\right\}^2 \\ = \;& E\sum_{i=1}^{n}\sum_{j=1}^{n}\{Y_i - E(Y_i|U_i))\}\{Y_j - E(Y_j|U_j)\}W_{n,i}(t,h_n)W_{n,j}(t,h_n) \end{aligned}$$

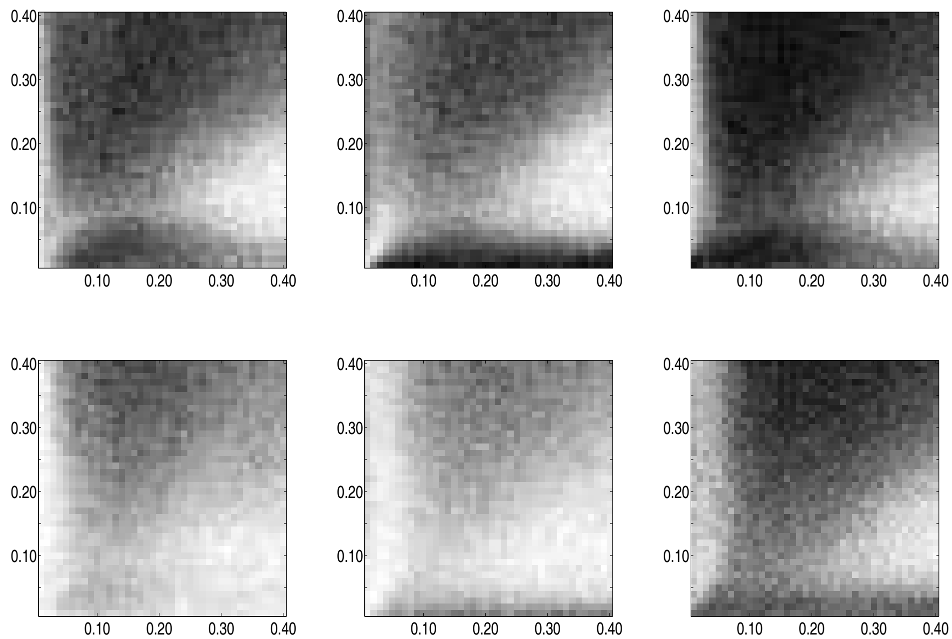

Figure 9. Grey-scale images of the P-values for Example 4 for comparing years 2002 with 2003, where in each frame the bandwidths corresponding to the estimate for the year 2002 are plotted along the horizontal axis, and those corresponding to the estimate for the year 2003 are plotted along the vertical axis. The top and the bottom rows correspond to the test statistics T_{n_1,n_2} and S_{n_1,n_2} respectively. The first two columns are images of bootstrapped P-values obtained by using the years 2002 and 2003 as the null hypothesis populations, respectively. The third column gives images of P-values obtained by the swap method.

$$
\begin{aligned}
&= \sum_{i=1}^{n} E\{Y_i - E(Y_i|U_i)\}^2 \, W_{n,i}^2(t,h_n) \\
&\leq K_0 \sum_{i=1}^{n} W_{n,i}^2(t,h_n) \longrightarrow 0 \text{ by assumption (D1) and (W1).}
\end{aligned}
$$

To show that the conditional bias goes to 0 almost surely, let $\epsilon > 0$ be given.

$$
\begin{aligned}
&E(\widehat{m}_n(t)) - m(t) \\
&\quad = E(Y_i|U_i)W_{n,i}(t,h_n) - m(t) \\
&\quad = \sum_{i=1}^{n}\{E(Y_i|U_i) - m(t)\}W_{n,i}(t,h_n) + m(t)\left\{\sum_{i=1}^{n} W_{n,i}(t,h_n) - 1\right\}.
\end{aligned}
$$

Note that the second term on the R.H.S goes to 0 by assumption (W2). Given $\frac{\epsilon}{4M} > 0$, $\exists\ A > 0$ and $N \in \mathbb{N}$ such that

$$\sum_{i=1}^{n} |W_{n,i}(t,h_n)| \mathbf{1}_{\left\{\left|\frac{t-F_n(X_i)}{h_n}\right|>A\right\}} \leq \frac{\epsilon}{4M}\ \forall\ n \geq N_0$$

by (W3). Now,

$$\begin{aligned}
&\left|\sum_{i=1}^{n}\{E(Y_i|U_i) - m(t)\}W_{n,i}(t,h_n)\right| \\
&\leq \left|\sum_{i=1}^{n}\{m(U_i) - m(t)\}W_{n,i}(t,h_n)\mathbf{1}_{\left\{\left|\frac{t-F_n(X_i)}{h_n}\right|\leq A\right\}}\right| \\
&\quad + \left|\sum_{i=1}^{n}\{m(U_i) - m(t)\}W_{n,i}(t,h_n)\mathbf{1}_{\left\{\left|\frac{t-F_n(X_i)}{h_n}\right|> A\right\}}\right| \\
&\leq \left|\sum_{i=1}^{n}\{m(U_i) - m(t)\}W_{n,i}(t,h_n)\mathbf{1}_{\left\{\left|\frac{t-F_n(X_i)}{h_n}\right|\leq A\right\}}\right| \\
&\quad +\epsilon/2\ \forall\ n \geq N_0 \text{ by assumption (W3)} \\
&\leq \sum_{i=1}^{n} |m(U_i) - m(t)||W_{n,i}(t,h_n)|\mathbf{1}_{\left\{\left|\frac{t-F_n(X_i)}{h_n}\right|\leq A\right\}} + \epsilon/2\ \forall\ n \geq N_0. \qquad (10)
\end{aligned}$$

Let $D_n = \sup_{t\in\mathbb{R}} |F_n(t) - F(t)|$ where F is the distribution function of X_1 and F_n is the empirical distribution function based on observations $X_1, X_2, \ldots, X_n$. Note that $D_n \geq |F(X_{(i)}) - F_n(X_{(i)})| = |U_{(i)} - \frac{i}{n+1}|$. By Glivenko Theorem, we know that $D_n \longrightarrow 0$ almost surely. Let D be the event $\{D_n \longrightarrow 0\}$. Note that $P(D) = 1$. Let us restrict to this event by intersecting other events with D.

Now, observe that $\left|\frac{t-i/(n+1)}{h_n}\right| \leq A$ and $|U_{(i)} - \frac{i}{n+1}| \leq D_n \Rightarrow |U_{(i)} - t| \leq Ah_n + D_n$. As $m(\cdot)$ is continuous, there exists $\delta > 0$ such that $|U_{(i)} - t| \leq \delta$ implies $|m(U_{(i)}) - m(t)| \leq \epsilon/2C$, where $\sum_{i=1}^{n} |W_{n,i}(t,h_n)| < C\ \forall\ n \geq 1$ by (W4). On the set D, $\exists\ N_1 \in \mathbb{N}$ such that $n \geq N_1 \Rightarrow Ah_n + D_n \leq \delta$. Thus, on the set $D\ \forall\ n \geq N_1$, we have $|m(U_i) - m(t)| \leq \epsilon/2C \Rightarrow \sum_{i=1}^{n} |m(U_i) - m(t)||W_{n,i}(t,h_n)|\mathbf{1}_{\left\{\left|\frac{t-F_n(X_i)}{h_n}\right|\leq A\right\}} \leq \epsilon/2$. Therefore, on the set D, $\forall\ n \geq \max\{N_1, N_0\}$, we have from equation (10), $|\sum_{i=1}^{n}\{E(Y_i|U_i) - m(t)\}W_{n,i}(t,h_n)| \leq \epsilon$. This proves that $|\sum_{i=1}^{n}\{m(U_i) - m(t)\}W_{n,i}(t,h_n)| \longrightarrow 0$ almost surely.

Now, using Chebyshev's inequality, we have

$$P(|\widehat{m}_n(t) - m(t)| \geq \epsilon | U_1, U_2, \ldots, U_n) \leq \frac{E(\{\widehat{m}_n(t) - m(t)\}^2 | U_1, U_2, \ldots, U_n)}{\epsilon^2} \xrightarrow{a.s.} 0 \text{ as } n \to \infty. \quad (11)$$

This implies that $\widehat{m}_n(t) \xrightarrow{P} m(t)$ as $n \to \infty$. □

PROOF OF COROLLARY 3.2. Let us first observe that

$$\begin{aligned}
& \left| \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{t - F_n(X_i)}{h_n} \right) - \int_0^1 \frac{1}{h_n} K\left(\frac{t-s}{h_n} \right) ds \right| \\
= & \left| \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{t - \frac{i}{n+1}}{h_n} \right) - \sum_{i=1}^{n} \int_{(i-1)/n}^{i/n} \frac{1}{h_n} K\left(\frac{t-s}{h_n} \right) ds \right| \\
= & \left| \frac{1}{nh_n} \sum_{i=1}^{n} \left\{ K\left(\frac{t - \frac{i}{n+1}}{h_n} \right) - K\left(\frac{t - s_i}{h_n} \right) \right\} \right|, \\
& \text{where } s_i \text{ is between } \frac{i-1}{n} \text{ and } \frac{i}{n} \ \forall\, i = 1, 2, \ldots, n \\
= & \left| \frac{1}{nh_n} \sum_{i=1}^{n} \left\{ K'(\xi_i) \left(\frac{s_i - \frac{i}{n+1}}{h_n} \right) \right\} \right|, \\
& \text{where } \xi_i \text{ is between } \frac{t - \frac{i}{n+1}}{h_n} \text{ and } \frac{t - s_i}{h_n} \ \forall\, i = 1, 2, \ldots, n \\
\leq & \frac{1}{nh_n} \sum_{i=1}^{n} \left| \frac{s_i - \frac{i}{n+1}}{h_n} \right| |K'(\xi_i)|. \qquad (12)
\end{aligned}$$

Assuming (C1) and (K2), and using (12), we now have

$$\frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{t - F_n(X_i)}{h_n} \right) = \int_0^1 \frac{1}{h_n} K\left(\frac{t-s}{h_n} \right) ds + O\left(\frac{1}{nh_n^2} \right).$$

Thus, we have $\sum_{i=1}^{n} \frac{1}{nh_n} K\left(\frac{t - F_n(X_i)}{h_n} \right) \to \int_0^1 \frac{1}{h_n} K\left(\frac{t-s}{h_n} \right) ds = 1$ as $n \to \infty$.

On the other hand, if we assume (C1) and (K1), we have

$$\frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{t - F_n(X_i)}{h_n} \right) = \int_0^1 \frac{1}{h_n} K\left(\frac{t-s}{h_n} \right) ds + O\left(\frac{1}{nh_n}\right)$$

using the fact that $K'(\xi_i) = 0$ if $|\xi_i| > M_0$ and there are $O(nh_n)$ nonzero terms in the sum in (12).

For the Nadaraya-Watson type estimator (W2) and (W4) are trivially true, while (W1) can be easily verified. Note also that for a kernel with compact support, (W3) is trivially satisfied. To verify (W3) for kernels with support on the entire real line, observe that

$$W_{n,i}(t,h_n) = \frac{K\left(\frac{t-F_n(X_i)}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{t-F_n(X_i)}{h_n}\right)} = \frac{\frac{1}{nh_n}K\left(\frac{t-F_n(X_i)}{h_n}\right)}{\sum_{i=1}^n \frac{1}{nh_n}K\left(\frac{t-F_n(X_i)}{h_n}\right)}, \ i = 1,2,\ldots,n,$$

where the denominator converges to 1. Using the fact that K integrates to 1, the tail integral converges to 0, which implies (W3).

Next consider the Priestley-Chao type estimator. Here (W1) follows trivially, and Equation (12) provides the proof of (W2) and (W3). Further, (W4) follows from that fact that a convergent sequence must be bounded.

In the case of Gasser-Muller type estimator, (W2) and (W4) are trivially true as in the case of Nadaraya-Watson type estimator. It is also easy to prove (W1) for this estimator. The verification of (W3) follows essentially the same lines of arguments as in the case of Nadaraya-Watson type estimator.

Finally, for the proof of the consistency of the local linear type estimator, see the arguments in Wand and Jones (1995) pp. 120-122 for the fixed equally spaced design model. The proof of our result follows from essentially same arguments. □

Proof of Theorem 3.3. Note that,

$$\begin{aligned}\widehat{m}_n(t)-m(t) &= \sum_{i=1}^n W_{n,i}(t,h_n)\{m(F(X_i))+\epsilon_i\} - m(t) \\ &= \sum_{i=1}^n W_{n,i}(t,h_n)\{m(F(X_i)) - m(t)\} \\ &\quad + \left\{\sum_{i=1}^n W_{n,i}(t,h_n) - 1\right\} m(t) + \sum_{i=1}^n W_{n,i}(t,h_n)\epsilon_i\end{aligned}$$

To find the limiting distribution of $\sum_{i=1}^n W_{n,i}(t,h_n)\epsilon_i$ let us define $Z_{n,i} = W_{n,(i)}(t,h_n)\epsilon_i$ for $i = 1,2,\ldots,n$, where $W_{n,(i)}(t,h_n)$ corresponds to the weight for the ith order-statistic of the X's. Let $S_n = \sum_{i=1}^n Z_{n,i}$. Note that $W_{n,(i)}(t,h_n)$ is just a constant which does not depend on the X's. Also notice that $S_n \stackrel{d}{=} \sum_{i=1}^n W_{n,i}(t,h_n)\epsilon_i$ due to the independence of the X's and the ϵ's (as we can re-arrange the ϵ's). We use the Lindeberg-Feller Central Limit Theorem to find the asymptotic distribution of S_n.

Note that $E(Z_{n,i}) = 0$ and $\sigma^2_{n,i} = Var(Z_{n,i}) = \sigma^2 W^2_{n,(i)}(t,h_n)$. Let $s_n^2 = \sum_{i=1}^n \sigma^2_{n,i} = \sigma^2 \sum_{i=1}^n W^2_{n,(i)}(t,h_n)$. For any $\eta > 0$,

$$\begin{aligned}
\sum_{i=1}^n \frac{1}{s_n^2} \int_{|Z_{n,i}|>\eta s_n} Z^2_{n,i}\ \mathrm{dP} &= \sum_{i=1}^n \frac{1}{s_n^2} \int_{\epsilon_i^2 > \frac{\eta^2 s_n^2}{W^2_{n,(i)}(t,h_n)}} W^2_{n,(i)}(t,h_n)\epsilon_i^2\ \mathrm{dP} \\
&\le \sum_{i=1}^n \frac{1}{s_n^2} \int_{\epsilon_i^2 > \frac{\eta^2 s_n^2}{\max_{1\le i\le n} W^2_{n,(i)}(t,h_n)}} W^2_{n,(i)}(t,h_n)\epsilon_i^2\ \mathrm{dP} \\
&\le \sigma^2 \int_{\epsilon_1^2 > \frac{\eta^2 s_n^2}{\max_{1\le i\le n} W^2_{n,i}(t,h_n)}} \epsilon_1^2\ \mathrm{dP} \ \longrightarrow 0 \text{ as } n\to\infty
\end{aligned}$$

by $(\tilde{W}3)$. Therefore by the Lindeberg-Feller Central Limit Theorem we have

$$\frac{\sum_{i=1}^n W_{n,i}(t,h_n)\epsilon_i}{\sigma\sqrt{\sum_{i=1}^n W^2_{n,i}(t,h_n)}} \xrightarrow{d} N(0,1) \text{ as } n\to\infty.$$

Also observe that we have by $(\tilde{W}1)$

$$\frac{\{\sum_{i=1}^n W_{n,i}(t,h_n) - 1\}m(t)}{\sigma\sqrt{\sum_{i=1}^n W^2_{n,i}(t,h_n)}} \longrightarrow 0 \text{ as } n\to\infty.$$

Next, we split $\sum_{i=1}^n W_{n,i}(t,h_n)\{m(F(X_i)) - m(t)\}$ as $\sum_{i=1}^n W_{n,i}(t,h_n)$ $\{m(F(X_i)) - m(F_n(X_i))\} + \sum_{i=1}^n W_{n,i}(t,h_n)\{m(F_n(X_i)) - m(t)\}$ and deal with the two terms separately. To simplify notations, let $A_n = \{i : |\frac{t-F_n(X_i)}{h_n}| \le A\}$. We simplify the first term as

$$\begin{aligned}
&\left|\frac{\sum_{i=1}^n W_{n,i}(t,h_n)\{m(F_n(X_i)) - m(F(X_i))\}}{\sigma\sqrt{\sum_{i=1}^n W^2_{n,i}(t,h_n)}}\right| \\
&= \left|\frac{\sum_{i\in A_n} W_{n,i}(t,h_n)m'(\xi_i)\{F_n(X_i) - F(X_i)\}}{\sigma\sqrt{\sum_{i=1}^n W^2_{n,i}(t,h_n)}}\right| \\
&\le \frac{1}{\sigma}\sqrt{\sum_{i\in A_n}\{m'(\xi_i)(F_n(X_i) - F(X_i))\}^2} \\
&\le \frac{K}{\sigma}\sqrt{\sum_{i\in A_n}\sup_{t\in\mathbb{R}}|F_n(t) - F(t)|^2} \\
&\le \frac{K}{\sigma}A\sqrt{2(n+1)}h_n^{1/2}\sup_{t\in\mathbb{R}}|F_n(t) - F(t)| \xrightarrow{P} 0 \text{ as } n\to\infty.
\end{aligned}$$

The second term can be simplified as

$$\left|\frac{\sum_{i=1}^{n} W_{n,i}(t,h_n)\{m(F_n(X_i)) - m(t)\}}{\sigma\sqrt{\sum_{i=1}^{n} W_{n,i}^2(t,h_n)}}\right|$$

$$= \left|\frac{\sum_{i\in A_n} W_{n,i}(t,h_n)m'(\xi_i)(F_n(X_i) - t)}{\sigma\sqrt{\sum_{i=1}^{n} W_{n,i}^2(t,h_n)}}\right| \text{ where } \xi_i \text{ is between } F_n(X_i) \text{ and } t$$

$$\leq \frac{1}{\sigma}\sqrt{\sum_{i\in A_n}\{m'(\xi_i)(F_n(X_i)) - t)\}^2} \;\leq\; \frac{K}{\sigma}\sqrt{\sum_{i\in A_n} A^2 h_n^2}$$

$$\leq \frac{K}{\sigma} A\sqrt{2(n+1)}h_n^{3/2} \longrightarrow 0 \text{ as } n \to \infty$$

by Cauchy-Schwarz inequality, $(\tilde{W}4)$ and noting that $|\{i : |\frac{t-F_n(X_i)}{h_n}| \leq A\}| \leq 2(n+1)Ah_n$. Combining all the three terms we have the required result. □

Proof of Corollary 3.4. As we have a compact kernel, $(\tilde{W}2)$ is satisfied by taking $A = M_0$. Also $(\tilde{W}4)$ holds by the assumption of the theorem. For the Nadaraya-Watson type estimator, $(\tilde{W}1)$ holds trivially as $\sum_{i=1}^{n} W_{n,i}(t,h_n) = 1$, $\forall\, n$. Note that LHS of $(\tilde{W}3)$ is $\sum_{i=1}^{n} K^2\left(\frac{t-F_n(X_i)}{h_n}\right)\Big/\max_{1\leq i\leq n} K^2\left(\frac{t-F_n(X_i)}{h_n}\right)$ which goes to infinity as $n \to \infty$ (since the kernel has a nonzero neighbourhood around 0 within which there will be infinitely many points as $n \to \infty$).

For the Gasser-Muller type weight function, $(\tilde{W}1)$ holds as $\sum_{i=1}^{n} W_{n,i}(t, h_n) = \int_{(t-1)/h_n}^{t/h_n} K(u)du = 1$ for large n (h_n should be small enough). $(\tilde{W}3)$ holds because of similar reasons as above. □

Acknowledgements. I am greatly indebted to my teacher, Probal Chaudhuri, whose guidance has been of invaluable help in developing the ideas presented here. I am grateful to Samsiddhi Bhattacharjee, Roger Koenker, Jayanta Kumar Pal, Dipankar Coondoo, Debasis Sengupta and the two referees for their helpful comments and suggestions. I would also like to thank A.K. Nag and the Reserve Bank of India for the data sets they provided.

References

Bhattacharya, P.K. and Muller, H,G. (1993). Asymptotics for nonparametric regression. *Sankhyā Ser. A*, **53**, 420-441.

Chaudhuri, P. and Marron, J.S. (1999). SiZer for exploration of structures in curves. *J. Amer. Statist. Assoc.*, **94**, 807-823.

CHAUDHURI, P. and MARRON, J.S. (2000). Scale space view of curve estimation. *Ann. Statist.*, **28**, 408-428.

DAVID, H.A. and NAGARAJA, H.N. (1998). Concomitants of order statistics. In *Handbook of Statistics* **16**, *Order Statistics: Thoery and Methods*, N. Balakrishnan and C.R. Rao, eds., 487-513, Elsevier, Amsterdam.

DAVID, H.A. and NAGARAJA, H.N. (2003). *Order Statistics* (3rd edition), Wiley, New York.

DELGADO, M.A. (1993). Testing the equality of nonparametric regression curves. *Statist. Probab. Lett.*, **17**, 199-204.

EFRON, B. and TIBSHIRANI, R.J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New York.

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modeling and Its Application*. Chapman and Hall, London.

HALL, P. and HART, J.D. (1990). Bootstrap test for difference between means in nonparametric regression. *J. Amer. Statist. Assoc.*, **85**, 1039-1049.

HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New York.

HALL, P. and HUBER, C. and SPECKMAN, P.L. (1997). Covariate-matched one-sided tests for the difference between functional means. *J. Amer. Statist. Assoc.*, **92**, 1074-1083.

HÄRDLE, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press, Cambridge.

HÄRDLE, W. and MARRON, J.S. (1990). Semiparametric comparsion of regression curves. *Ann. Statist.*, **18**, 63-89.

KING, E.C., HART, J.D. and WEHRLY, T.E. (1991). Testing the equality of two regression curves using linear smoothers. *Statist. Probab. Lett.*, **12**, 239-247.

KULSEKERA, K.B. (1995). Comparison of regresion curves using quasi residuals. *J. Amer. Statist. Assoc.*, **90**, 1085-1093.

KULSEKERA, K.B. and WANG, J. (1997). Smoothing parameter selection for power optimality in testing of regression curves. *J. Amer. Statist. Assoc.*, **92**, 500-511.

MAHALANOBIS, P.C. (1960). A method for fractile graphical analysis. *Econometrica*, **28**, 325-351.

MAHALANOBIS, P.C. (1988). *Fractile Graphical Analysis*, P.K. Bose, ed., Statistical Publishing Society, Calcutta.

MITROFANOVA, N.M. (1961). On some problems of fractile graphical analysis. *Sankhyā Ser. A*, **23**, 145-154.

MÜLLER, H.G. (1988). *Nonparametric Regression Analysis of Longitudinal Data*. Springer-Verlag, Berlin.

MUNK, A. and DETTE, H. (1998). Nonparametric comparison of several regression functions: exact and asymptotic theory. *Ann. Statist.*, **26**, 2339-2368.

NADARAYA, E.A. (1964). On estimating regression. *Theor. Probability Appl.*, **9**, 141-142.

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

PARTHASARATHY, K.R. and BHATTACHARYYA, P.K. (1961). Some limit theorems in regression theory. *Sankhyā Ser. A*, **23**, 91-102.

PRIESTLEY, M.B. and CHAO, M.T. (1972). Nonparametric function fitting. *J. Roy. Statist. Soc. Ser. B*, **34**, 385-392.

Rice, J.A. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.*, **12**, 1215-1230.

Sethuraman, J. (1961). Some limit distributions connected with fractile graphical analysis. *Sankhyā Ser. A*, **23**, 79-90.

Simonoff, J.S. (1996). *Smoothing Methods in Statistics*. Springer-Verlag, New York.

Stone, C.J. (1977). Consistent nonparametric regression. *Ann. Statist.*, **5**, 505-545.

Takeuchi, K. (1961). On some properties of error area in the fractile graph method. *Sankhyā Ser. A*, **23**, 65-78.

Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*. Chapman and Hall, London.

Watson, C.S. (1964). Smooth regression analysis. *Sankhyā Ser. A*, **26**, 359-372.

Yang, S.S (1977). General distribution theory of the concomitants of order statistics. *Ann. Statist.*, **5**, 996-1002.

Young, S.G. and Bowman, A.W. (1995). Nonparametric analysis of covariance. *Biometrics*, **51**, 920-931.

Bodhisattva Sen
Statistics Department
University of Michigan
Ann Arbor, MI 48109-1092
E-mail: bodhi@umich.edu

Paper received: August 2004; revised May 2005.